MANAGEMENT ET INFORMATIQUE

collection dirigée par Nicolas Manson

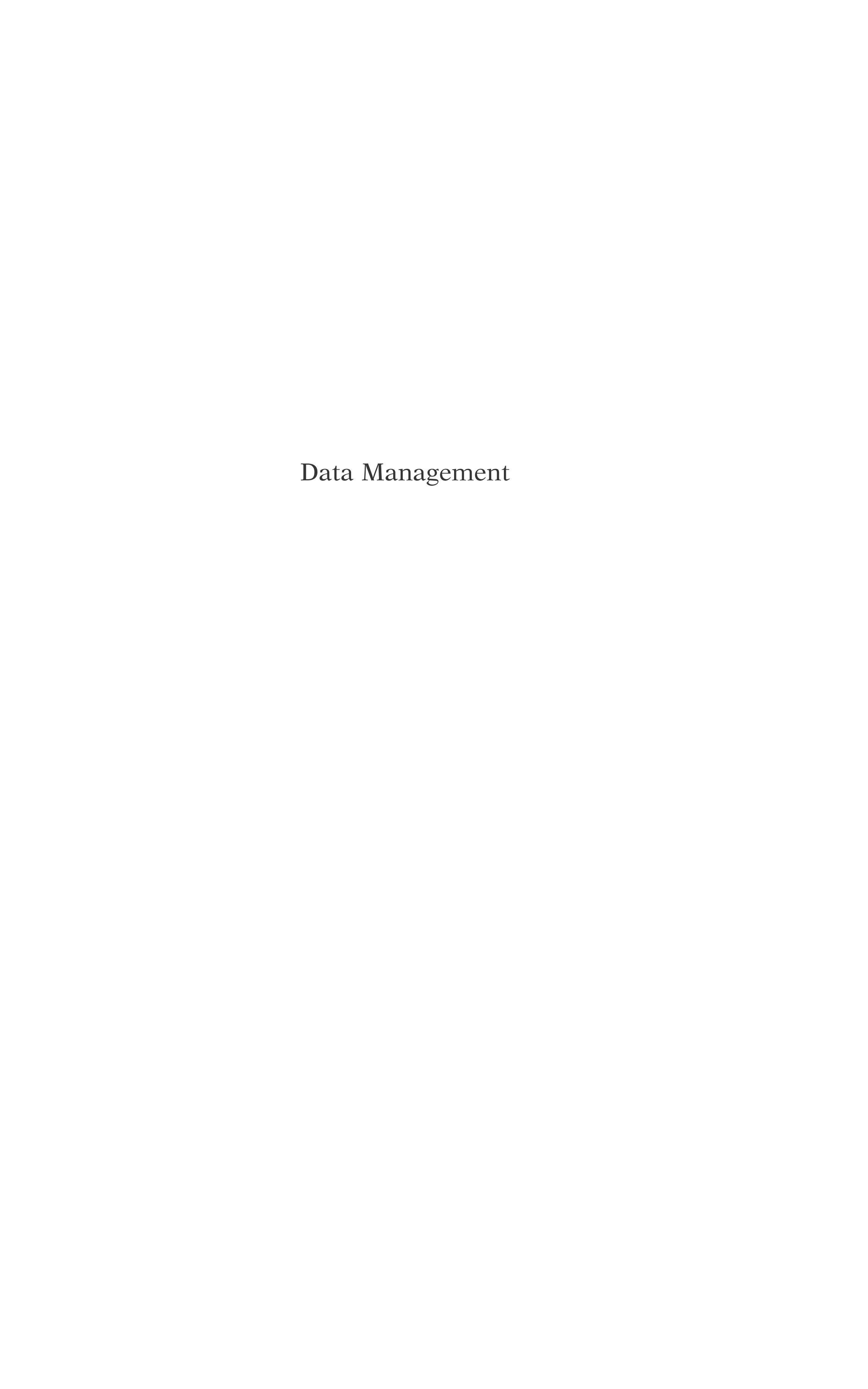
Data Management

qualité des données et compétitivité

Christophe Brasseur



Lavoisier



Data Management

A Christine, Arthur et Capucine

LAVOISIER
11, rue Lavoisier
75008 Paris

www.hermes-science.com
www.lavoisier.fr

ISBN 2-7462-1210-2

Data Management

qualité des données et compétitivité

Christophe Brasseur



COLLECTIONS SOUS LA DIRECTION DE NICOLAS MANSON

Collection Management et Informatique

Collection Etudes et Logiciels Informatiques

Collection Nouvelles Technologies Informatiques

Collection Synthèses Informatiques CNAM

La liste des titres de chaque collection se trouve en fin d'ouvrage.

Table des matières

INTRODUCTION

Qui est concerné par la qualité des données ?

Plus que jamais, les entreprises sont confrontées à des marchés de plus en plus exigeants, réclamant une réponse immédiate et adaptée à des besoins évoluant en permanence. Face à cette situation, tous les secteurs d'activité, de l'automobile aux télécommunications, en passant par la banque ou les services publics, s'efforcent d'adapter et de renouveler régulièrement leurs produits et services pour coller au mieux à la demande.

De la compréhension du marché à la livraison d'un produit ou d'un service, les entreprises doivent orchestrer la gestion de leur activité de façon optimale. Pour cela, elles ont généralement organisé et mis en place des procédures de gestion, automatisées ou non.

Dans ce contexte, la maîtrise de l'information est devenue indispensable et le système d'information qui accompagne l'intégration et la fluidité des processus joue un rôle essentiel. Toutes les fonctions de l'entreprise (comptabilité, contrôle de gestion, marketing, vente, production, achats, ressources humaines, etc., mais aussi qualité, maintenance ou recherche) sont désormais couvertes par le système d'information.

Très clairement, tout le monde est concerné par des données de qualité, puisque celles-ci supportent la plupart des flux de l'entreprise. Les seules personnes qui n'ont pas besoin de s'en préoccuper sont celles qui n'en créent pas et n'en utilisent jamais. Or dans un monde irrigué d'informations et exigeant de plus en plus de réactivité, qui peut prétendre vivre à l'écart des données ?

Les *dirigeants* : ils sont concernés au plus haut point, puisque les tâches de définition de la stratégie, de prise de décisions, de pilotage de l'entreprise nécessitent plus que jamais d'informations de grande qualité. Les dirigeants sont conscients qu'avec des données de qualité, leur entreprise est plus forte, plus réactive, plus flexible donc plus compétitive. Aussi, mieux connaître ses clients pour les satisfaire davantage est l'une de leurs préoccupations majeures. La qualité des informations contenues dans la base clients est à cet égard fondamental.

L'avenir de l'entreprise ne peut donc pas se contenter d'un système d'information contenant des données de qualité médiocre. Les dirigeants doivent donc veiller à ce que le patrimoine informationnel soit à la hauteur des ambitions de leur société. Mais il ne suffit pas pour autant de déléguer cette mission à la Direction des systèmes d'information ou à la Direction de l'informatique. Pour être efficaces, la communication et les actions visant à obtenir une qualité d'information correcte, doivent être portées par les dirigeants eux-mêmes.

En d'autres termes, puisque les dirigeants sont largement concernés par la qualité des données, ils doivent être les premiers à promouvoir une culture d'entreprise laissant une place importante à la qualité de l'information et des données. Le jeu en vaut la chandelle et l'investissement dans une telle politique est clairement rentable.

Les *responsables opérationnels* : en charge d'un centre de profit, d'une unité de production ou d'un service transverse (marketing, achats, comptabilité, ressources humaines…), ces

managers opérationnels gèrent au quotidien une grande quantité de tâches et de problèmes variés. A l'interface de la direction générale et des tâches d'exécution, ils sont utilisateurs d'un grand nombre d'informations et, en conséquence, la qualité des données impacte directement leur job. Ils doivent en effet s'assurer à la fois de la bonne marche des opérations qui leur sont confiées et être capables de remonter à la direction générale des éléments pertinents et synthétiques.

Ainsi, les responsables marketing et commerciaux sont concernés par la qualité des données car celles-ci représentent quasiment toute la connaissance des clients, des fournisseurs, des concurrents, du marché.

Reconnaissons toutefois que dans le domaine commercial, tout n'est pas décrit sous forme de données. Par exemple, les commerciaux ont en tête des éléments comme le comportement ou les motivations de tel ou tel prospect ou client. Hormis ces exceptions, qui n'en sont pas moins précieuses, la qualité des données procure un avantage compétitif évident.

Les *collaborateurs opérationnels* : au quotidien, le travail de la grande majorité des employés est largement supporté par des données. Ainsi, un approvisionneur doit connaître pour chaque article un niveau de stock fiable, sous peine de subir des ruptures de stock intempestives ou au contraire de commander à tort de la marchandise. Autre exemple, un technicien de maintenance doit pouvoir connaître la liste des dernières interventions effectuées sur une machine donnée, ainsi que les pièces changées. Si ces informations ne sont pas fiables, l'entretien de la machine risque de ne pas être effectué au bon moment et le taux de panne ira en augmentant. Ou encore, une assistante commerciale doit pouvoir confirmer une commande et indiquer une date de livraison fiable à un client. Des erreurs répétées dans les dates de livraison feront fuir à juste titre les clients mécontents.

Ces exemples pourraient être multipliés à l'infini et il n'est pas exagéré d'affirmer que tous les métiers opérationnels de l'entreprise nécessitent à un moment ou un autre la création ou l'utilisation de données correctes.

Notons que l'information est souvent partagée entre différents acteurs. En effet, la plupart des personnes utilisent des données fournies par d'autres et créent elles-mêmes des données qui alimenteront le travail d'autres collègues. D'où l'importance de veiller à maintenir la qualité des données tout au long de la chaîne d'information.

Les *responsables informatiques* : ils se préoccupent en permanence de la qualité de l'information qui circule dans le système d'information. Si un problème de données survient, c'est naturellement vers leurs équipes que l'on se tourne.

Mais contrairement aux idées reçues, les responsables informatiques ne peuvent pas et ne doivent pas être responsables de la qualité de toutes les données de l'entreprise. Tout simplement parce qu'ils ne contrôlent pas tout et ne sont généralement pas à la source de la création des données. De nombreux managers informatiques se considèrent à tort responsables de toutes les données de l'entreprise, y compris de celles qui sont créées chaque jour par les utilisateurs.

Le rôle le plus important des responsables informatiques et de leurs équipes à l'égard des données est sans aucun doute l'animation et la promotion du rôle-clé de la qualité de l'information. Faire de la prévention, former, conseiller, aider, accompagner les utilisateurs font partie intégrante de leur job.

Le Data Management

Malgré l'évolution rapide des technologies et le développement continu des systèmes d'information au sein des entreprises, la qualité des données manipulées demeure souvent

le point faible. Les logiciels collent de mieux en mieux aux processus métiers de l'entreprise, mais le contenu est souvent négligé. Les systèmes regorgent de données obsolètes, incorrectes ou incomplètes, sans parler des doublons (une même réalité enregistrée plusieurs fois) qui perturbent l'utilisation opérationnelle et polluent les analyses. Comment tirer profit du système d'information si le contenu n'est pas fiable et de qualité médiocre ?

La vraie difficulté est que la mauvaise qualité des données ne se détecte pas facilement. Ce sont souvent des incidents ou des anomalies dans le travail opérationnel qui révèlent ici et là des inconsistances portant sur les données. Par exemple, un client non livré révélera que la base de données clients n'est pas à jour ou que la commande a été saisie sur un enregistrement client ayant une adresse incorrecte.

Mais il y a plus grave. Avec des données incorrectes ou incomplètes, les indicateurs d'activité peuvent s'avérer éloignés de la réalité et conduire à des décisions inadaptées. Aussi, avec l'expansion d'internet, l'image même de l'entreprise peut rapidement être détériorée, en particulier si des informations publiées comme les prix ne sont pas actualisées.

Il y a enfin souvent beaucoup trop de données maintenues qui ne servent finalement à rien. Certains dirigeants se plaignent que leur système d'information délivre une quantité d'informations détaillées d'un faible intérêt et ne fournit pas les quatre ou cinq données vraiment utiles à leur job.

Les bonnes pratiques de Data Management, c'est-à-dire de gestion des données sont incontestablement la réponse adaptée à ce type de problèmes. Basé sur des principes clairs, privilégiant les mesures préventives et le pragmatisme, le Data Management a pour objectif d'améliorer la qualité des données et de rendre l'entreprise plus performante et compétitive.

Objectifs de l'ouvrage

Ce livre volontairement synthétique se propose donc d'expliquer les enjeux de la qualité des données et de définir les bonnes pratiques de Data Management (gestion des données).

Data Management s'adresse principalement aux dirigeants d'entreprises, aux managers opérationnels et aux professionnels des systèmes d'information (responsables informatiques, chefs de projets, consultants, ingénieurs). Il s'adresse également aux étudiants d'écoles de commerce, d'ingénieurs et des universités soucieux de comprendre les problématiques de qualité de l'information.

La première partie introduit le rôle-clé des données dans les systèmes d'information de gestion. Elle décrit aussi quelques grandes tendances actuelles en matière d'applications (ERP, e-Business, Business Intelligence, CRM…) pour lesquelles la qualité des données est essentielle.

La deuxième partie explique les enjeux de la qualité de l'information et des données, en mettant clairement en évidence les impacts sur la performance et la compétitivité des entreprises.

La troisième partie décrit très concrètement les méthodes permettant d'améliorer la qualité des données. De nombreuses recommandations pratiques enrichissent les méthodes décrites.

Enfin, la dernière partie est entièrement consacrée à la reprise des données. Si la migration d'une ancienne application vers une nouvelle est relativement fréquente, force est de constater qu'une majorité de problèmes survenant au démarrage d'un nouveau système proviennent de données incorrectes ou incomplètes. Il est donc absolument indispensable d'initialiser tout nouveau système avec des données de qualité.

PREMIÈRE PARTIE

Systèmes d'information et données

CHAPITRE 1

L'information, matière première vitale de l'entreprise

1.1. L'évolution des systèmes d'information

Les vagues technologiques qui se sont succédé au cours de ces trente dernières années ont joué un rôle déterminant dans l'amélioration des systèmes d'information.

Après les grands systèmes centralisés qui ont fait d'IBM un leader dès les années 1960, sont apparus les systèmes « client-serveurs ». Ceux-ci laissent davantage d'autonomie aux utilisateurs finaux connectés à des serveurs *via* leur PC, véritable révolution des années 1980.

Dans les années 1990, les progiciels de gestion intégrés construits sur des bases de données relationnelles ont été adoptés par une majorité de grandes entreprises. Ces produits logiciels préfabriqués appelés aussi ERP (*Entreprise Resource Planning*) permettent de gérer toutes les fonctions de l'entreprise en assurant une parfaite intégration entre elles.

La fin des années 1990 est marquée par la révolution internet. Les excès de cette période qui se sont traduits par l'éclatement

de la bulle financière ne doivent pas masquer l'énorme bouleversement et la tendance de fond apportée par le net : partage et diffusion mondiale de l'information, explosion du commerce électronique.

Aujourd'hui la technologie sans fil c'est-à-dire le Wi-Fi (*Wireless Fidelity*) est prometteuse, car elle permet l'accès et l'échange de données sans la contrainte d'un poste fixe ou de câbles.

Les progrès constants réalisés, d'une part, dans les capacités de stockage des données, d'autre part, dans la puissance des processeurs, ont largement contribué à l'essor des systèmes d'information au sein des organisations. Aussi, la place croissante accordée aux directions des systèmes d'information au cours des trente dernières années témoigne du rôle stratégique de l'information. Ainsi la terminologie a évolué : la Direction de l'informatique a laissé place à la Direction des systèmes d'information. Dans les pays anglo-saxons, le CIO (*Chief Information Officer*) fait partie du comité de direction au même titre que le directeur du marketing ou le directeur financier.

Les budgets consacrés à l 'évolution des systèmes d'information sont évidemment liés à la santé des entreprises et d'une manière générale à la conjoncture économique. Mais, globalement, la tendance de l'investissement dans les technologies de l'information depuis une trentaine d'années est à la hausse. La bataille entre entreprises concurrentes se joue de plus en plus sur la qualité du système d'information.

1.2. Information et données

Même si l'on parle encore d'informatique, l'évolution de terminologie traduit une préoccupation fondamentale : optimiser la gestion de l'entreprise par la maîtrise de l'information. A l'inverse du système informatique qui est un ensemble de

matériels et de logiciels, le système d'information n'est pas dépendant d'une plate-forme matérielle.

Il est très difficile de définir le mot *information* car il revêt un sens très large et se retrouve dans différents domaines : presse, télévision, droit, économie, gestion… ce qui en fait une notion abstraite. Le dictionnaire indique : « action d'informer ou de s'informer ; renseignement quelconque » et confirme ainsi le caractère très étendu de ce mot.

Dans le contexte de la gestion, l'information est un ensemble d'éléments reflétant une réalité économique ou physique, susceptibles d'apporter de la connaissance utile à l'exercice de l'activité de l'entreprise. Tous les acteurs de l'entreprise, de l'ouvrier au dirigeant, ont besoin d'informations pour exercer leur métier dans les meilleures conditions. C'est clairement une matière première vitale nécessaire au fonctionnement de l'entreprise.

Pratiquement, l'information est déduite d'un ensemble de données. Les données sont donc des éléments porteurs d'information. Les deux mots ne sont pas synonymes, bien qu'ils soient souvent utilisés l'un pour l'autre. Les données peuvent aussi bien être définies par des individus que générées par le système d'information. Ainsi, la date de création d'une commande, le chiffre d'affaires d'une entité, la quantité en stock d'un article sont des exemples de données.

1.3. Les données : pour faire quoi et pour qui ?

La concurrence de plus en plus vive entre sociétés impose désormais d'être très vigilant quant à la qualité de l'information, car c'est un gage de compétitivité.

L'information dans l'entreprise est variée et dépend des besoins des différentes fonctions. Tous les métiers de l'entreprise ne peuvent être exercés avec efficacité que si

l'information adéquate est disponible au bon moment, avec le niveau de précision adapté. Certaines informations doivent par ailleurs être disponibles pour des raisons légales et fiscales. On peut globalement distinguer :

– les contraintes légales,

– les besoins opérationnels,

– les besoins d'analyse et d'aide à la décision.

1.3.1. *Les contraintes légales*

1.3.1.1. *La publication des comptes*

Dans un grand nombre de pays, les entreprises sont tenues de publier leurs comptes chaque année.

En France, les entreprises ont pour obligation de produire un bilan et un compte de résultat. Les entreprises cotées en Bourse doivent produire en plus un tableau des flux de trésorerie. A compter de 2005, les sociétés européennes cotées doivent établir leurs états financiers consolidés trimestriellement selon le référentiel IFRS (*International Financial Reporting Standards*).

1.3.1.1.1. Le bilan

Rappelons que le bilan est une photographie de l'entreprise à un instant *t*. Le bilan au dernier jour de l'exercice doit être publié et figure dans le rapport annuel des entreprises. Il comporte trois parties :

– l'actif : tout ce qui constitue les avoirs de l'entreprise (terrains, immeubles, bureaux, machines…),

– les dettes : les sommes enregistrées en dettes correspondent aux montants que l'entreprise devra décaisser (fournisseurs, personnel, banques…),

– les fonds propres : ils correspondent à la différence entre ce que l'entreprise a et ce qu'elle doit. On a donc l'égalité : fonds propres = actif – dettes. Le passif étant la somme fonds propres + dettes, on a donc l'égalité actif = passif.

1.3.1.1.2. Le compte de résultat

A l'inverse du bilan, le compte de résultat mesure les variations de recettes et de dépenses entre deux dates. Il traduit l'évolution des fonds propres de l'entreprise au cours de la période écoulée. On distingue donc :

– les produits : ce sont une source d'augmentation de richesse ; le chiffre d'affaires induit par les ventes est donc un produit ;

– les charges : ce sont une source de diminution de richesse ; les achats de matières premières ou les salaires versés au personnel sont des charges.

La différence entre produits et charges correspond au résultat avant impôts. Une fois amputé des impôts, on obtient le résultat net. En pratique, les produits et charges sont classées selon trois parties : la partie exploitation résultant de l'activité quotidienne de l'entreprise, la partie finance traitant des produits et charges financières, la partie exceptionnelle pour les opérations ne faisant pas partie de l'exploitation courante. Et on obtient donc selon la partie, un résultat d'exploitation, un résultat financier et un résultat exceptionnel. La somme de ces résultats permet d'aboutir au résultat avant impôt.

1.3.1.1.3. Le tableau de flux de trésorerie

Le tableau de flux permet d'apprécier la variation de trésorerie au cours de l'exercice.

L'ensemble des flux est présenté en trois parties :

– les flux d'exploitation, représentant les flux de l'entreprise avec les partenaires commerciaux (clients, fournisseurs, personnel…) ;

– les flux de financement représentant les flux de l'entreprise avec les partenaires financiers (banques, actionnaires…) ;

– le flux d'investissement, représentant les flux de l'entreprise avec ses fournisseurs d'immobilisations.

Pour chacun de ces flux, toute augmentation d'actif (emploi) et toute diminution ou dette (ressource) est chiffrée. La trésorerie à la fin de l'exercice se déduit de l'ensemble de ces flux.

1.3.1.2. *Le système d'information comptable*

Pour pouvoir être publiés, ces documents nécessitent un système d'information comptable où de nombreuses rubriques présentées précédemment de façon très synthétique, doivent être alimentées par des données fiables.

Les récentes normes IAS/IFRS ont pour but d'harmoniser les pratiques en matière de publication des résultats. Elles imposent en particulier de diffuser une information financière plus économique, dégagée de contraintes fiscales ou juridiques. Ainsi les bilans doivent être plus précis puisque les chiffres doivent désormais être ventilés par branche d'activité et par zone géographique. On peut dire globalement que les normes IFRS apportent davantage de transparence et que leur orientation est majoritairement tournée vers les investisseurs.

La comptabilité est l'une des premières fonctions de l'entreprise a avoir été informatisée et, en conséquence, les logiciels sont souvent très fiables et s'appuient sur les normes comptables en vigueur. Les anomalies rencontrées dans ce domaine résultent souvent d'une mauvaise utilisation de l'outil ou d'erreurs de saisie de montants.

Soulignons que d'autres domaines que la comptabilité sont aussi soumis à des obligations. Ceci a un impact direct sur la gestion des données. Ainsi, la sécurité, l'environnement, la santé nécessitent de produire des documents attestant que le produit ou le service délivré est conforme à la législation.

1.3.2. *Les besoins opérationnels*

La plupart des acteurs de l'entreprise ont besoin d'informations et manipulent des données pour exercer leur métier. Aussi, internet permet de diffuser plus facilement des informations à l'extérieur de l'entreprise.

Prenons quelques exemples :

– un assistant commercial doit pouvoir saisir des commandes clients et être capable de communiquer une date de livraison fiable à chaque client ;

– un acheteur doit pouvoir suivre le détail des achats effectués auprès de chaque fournisseur ;

– un responsable des stocks doit être capable de suivre le niveau des stocks et saisir des entrées ou des sorties d'articles ;

– un comptable doit pouvoir saisir des factures fournisseurs ;

– un responsable des ventes doit être capable de lister les clients stratégiques et obtenir le chiffre d'affaires réalisé ;

– un client ou prospect doit pouvoir consulter les prix d'un produit sur internet.

Remarquons que des données produites pour des besoins légaux, comme le compte de résultat, peuvent aussi être classées dans la catégorie des besoins opérationnels. Au-delà de l'aspect légal, il est évident que l'entreprise doit maîtriser ses comptes et savoir si elle gagne ou perd de l'argent.

Les besoins opérationnels nécessitent généralement :

– la saisie de données (création ou modification),
– la consultation de données,
– l'édition de documents (formulaires, listes),
– la diffusion de données.

Soulignons qu'un nombre non négligeable de données sont partagées par plusieurs métiers. Par exemple un article matière première intéresse un acheteur, mais aussi un responsable

fabrication, un responsable magasin ou encore un contrôleur de gestion. Certes, chacun d'eux a une vision différente de l'article en question et s'intéresse aux informations propres à son activité, mais il y a des informations communes.

La notion d'information partagée est extrêmement importante et il faut veiller à organiser les données de façon non redondante. C'est ce qui fait la force des progiciels de gestion intégrés (voir section 2.1). Ces progiciels sont construits sur des bases de données relationnelles et l'information n'est saisie qu'une seule fois à la source. Cela n'est cependant pas suffisant pour garantir des données de qualité.

1.3.3. *Les besoins d'analyse et d'aide à la décision*

Les managers de l'entreprise ont besoin d'informations synthétiques et pertinentes appelées indicateurs pour analyser l'activité de l'entreprise et prendre les meilleures décisions.

Les outils de pilotage répondent à ces besoins. Les données produites par ces outils sont stockées dans des bases de données dédiées appelées entrepôts de données (*Datawarehouse*). Le volume de données gérées est dans la plupart des cas considérable.

Les indicateurs sont généralement construits selon un ou plusieurs axes d'analyse. Par exemple, l'analyse des achats de matières premières peut se faire par famille, par article, par fournisseur, par mois. La définition des indicateurs et des axes d'analyse est donc une étape-clé et il faut avant tout identifier avec précision ce que l'on veut mesurer.

Soulignons que les informations de pilotage sont produites principalement à partir de données issues du système opérationnel. Par exemple, le chiffre d'affaires d'une entité pour un mois donné sera obtenu à partir des commandes enregistrées par cette entité sur la période considérée. D'où l'importance des

données saisies au niveau opérationnel. Le système de pilotage peut aussi être alimenté dans une moindre mesure à partir de données externes.

On peut schématiquement représenter les flux d'informations entre système opérationnel, système de pilotage et système externe de la façon suivante :

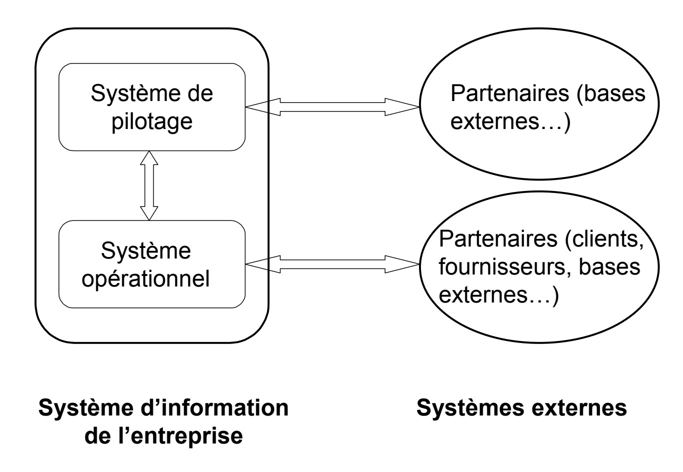


Figure 1.1. *Flux entre systèmes d'information*

1.4. Différents types de données

Si l'univers de la gestion utilise majoritairement les données sous forme texte, l'arrivée du multimédia et d'internet permettent d'exploiter d'autres formats de données. Cette section présente de façon synthétique les différents types de données manipulés par les ordinateurs.

1.4.1. *Le texte*

Dès les débuts de l'informatique, c'est naturellement le format texte qui a été utilisé. L'ordinateur ne manipulant que des chiffres 0 ou 1, l'idée était de représenter chaque caractère par une suite de 0 ou de 1.

Le code ASCII (*American Standard Code for Information Interchange*) a été la première norme établie en 1963 pour associer chaque caractère à une suite de 0 et de 1. Ce premier standard permettait de coder 128 caractères différents en associant à chacun d'eux une série de 7 bits (série de 7 chiffres prenant chacun 0 ou 1 comme valeur). Cette version purement anglo-saxonne ne pouvait cependant pas être utilisée par d'autres langues car elle n'intégrait aucun caractère accentué.

C'est la raison pour laquelle cette version a donné lieu au code ASCII étendu comprenant 8 bits et permettant de gérer 256 caractères. L'ASCII étendu inclut ainsi les caractères accentués d'autres langues (latin, arabe, grec, cyrillique…).

Plus récemment, la norme Unicode, qui permet de gérer les caractères de toutes les langues a étendu l'ASCII en utilisant 32 bits.

Pour l'utilisateur, cette codification interne sous forme de 0 et de 1 est complètement transparente et c'est bien du texte qu'il manipule. Par exemple, les données d'un fichier client : nom, adresse, code postal… apparaissent sous format texte (et heureusement !).

1.4.2. *Les dessins vectoriels*

Les possibilités de calcul des ordinateurs ont très tôt conduit à rechercher la représentation d'objets géométriques pour faciliter le travail de conception dans l'industrie et la construction. Les dessins vectoriels sont des objets géométriques dont la forme et la position sont déterminées par des formules mathématiques.

A partir des années 1980, les outils de DAO (dessin assisté par ordinateur) et de CAO (conception assistée par ordinateur) ont fait leur apparition en même temps que les stations de travail pour manipuler des dessins vectoriels. Les stations de travail

sont des ordinateurs dotés de puissantes possibilités de calculs et de visualisation. La puissance de l'ordinateur permet ainsi d'effectuer des calculs (par exemple la résistance d'un pont) et de la simulation (par exemple la déformation d'une automobile lors d'un choc à une certaine vitesse).

Les premières stations de travail étaient chères, de l'ordre de plusieurs centaines de milliers de francs voire plusieurs millions pour des systèmes complets. Seules les grandes entreprises pouvaient investir dans ces techniques. La chute des prix des ordinateurs a considérablement élargi le nombre d'utilisateurs de ces outils qui existent désormais sur PC.

Le nombre considérable de données nécessaires à ces applications entraîne généralement l'utilisation d'un SGDT (système de gestion des données techniques) comportant une base de données dédiée. Outre le domaine de l'industrie, le secteur des jeux vidéo utilise aujourd'hui largement les dessins vectoriels.

1.4.3. *L'image matricielle*

Une image matricielle est une succession de points colorés appelés pixels. Elle est obtenue en utilisant un appareil photo numérique ou en scannant un document. Le nombre de pixels et le nombre de nuances de couleurs influent directement sur la qualité de l'image restituée.

Standard	Nombre de points	Utilisation
VGA	640 X 480	Personnelle, bureau. De moins en moins utilisé
SVGA	800 X 600	Personnelle, bureau. Très utilisé
XGA	1024 X 1200	Personnelle, bureau. Très utilisé
SXGA	1280 X 1024	Professionnels de l'image, communication.
UXGA	1600 X 1200	Professionnels de l'image, communication.

Tableau 1.1. *Standards d'affichage des écrans d'ordinateurs*

Le tableau 1.1 indique les standards actuels d'affichage des écrans d'ordinateurs.

Les possibilités de nuances des couleurs dépendent du nombre de bits associés à chaque pixel. Si l'on ne dispose que d'un bit par pixel, on aura le choix entre 2 couleurs. 1 octet (8 bits) permettra d'obtenir 256 couleurs. Les codages les plus utilisés sont : 8 bits, 16 bits et 32 bits (ce dernier codage étant dit « couleurs vraies »).

Les images matricielles prennent beaucoup de place et les techniques de compression de fichiers permettent d'éviter cet inconvénient (formats Gif, Png, Jpeg).

La plupart des sites internet utilisent les images matricielles pour joindre des photos au texte.

1.4.4. *La vidéo numérique*

La vidéo est une succession d'images cadencées de façon à donner une impression de continuité (25 images par seconde). La technique consiste à utiliser des images matricielles qui se succèdent et l'on comprend aisément que la taille des fichiers est problématique.

Fort heureusement, les techniques de compression permettent de diminuer nettement la taille des fichiers. Ainsi, l'organisme MPEG (*Motion Picture Experts Group*) a établi un certain nombre de standards :

– MPEG-1 (1992) : de faible qualité ;

– MPEG-2 (1994) : de bonne qualité utilisé pour les DVD et la télévision numérique ;

– MPEG-4 (1999) : très compressé et utilisé par certains sites internet ou encore la téléphonie mobile ; il permet de faire tenir un film court sur 700 mégaoctets ;

– MPEG-7 (2001) : format de description des données audio et vidéo ; pas encore retenu par les industriels car techniquement complexe, il doit permettre d'indexer les contenus multimédias pour les retrouver plus facilement dans une base de données.

En entreprise, la vidéo numérique est principalement utilisée dans les domaines de la communication et de la formation. Le secteur de l'audiovisuel (télévision, cinéma) utilise de plus en plus les techniques numériques.

1.4.5. *Le son*

Le son est une vibration de l'air, c'est-à-dire une suite de surpressions et de dépressions de l'air par rapport à la pression atmosphérique.

Les haut-parleurs d'une enceinte acoustique fonctionnent selon ce principe. Le haut-parleur est constitué d'une membrane reliée à un électro-aimant, qui, suivant les sollicitations d'un courant électrique, va aller en avant et en arrière très rapidement, provoquant une vibration de l'air situé devant lui, délivrant ainsi du son.

Les caractéristiques du son sont les suivantes :

– la fréquence : c'est le nombre d'oscillations de pression par seconde ; elle se mesure en Herz ; plus elle est élevée, plus le son est aigu ; l'oreille permet de distinguer les fréquences de 20 Hz (graves) à 20 000 Hz (aigus) ;

– l'intensité : c'est la force du son qui se mesure en décibels ; à 0 db, c'est le silence ; au-delà de 130, l'intensité est difficilement supportable.

1.4.5.1. *La numérisation du son*

Pour être exploité par un ordinateur, il y a besoin de convertir le son en valeurs numériques : c'est la numérisation. Un son est

représenté informatiquement par plusieurs paramètres, décrits ci-dessous.

1.4.5.1.1. La fréquence d'échantillonnage

L'échantillonnage consiste à relever des petits échantillons de son à des intervalles de temps précis.

Autrement dit, il s'agit de découper un signal électrique pour en avoir une représentation numérique.

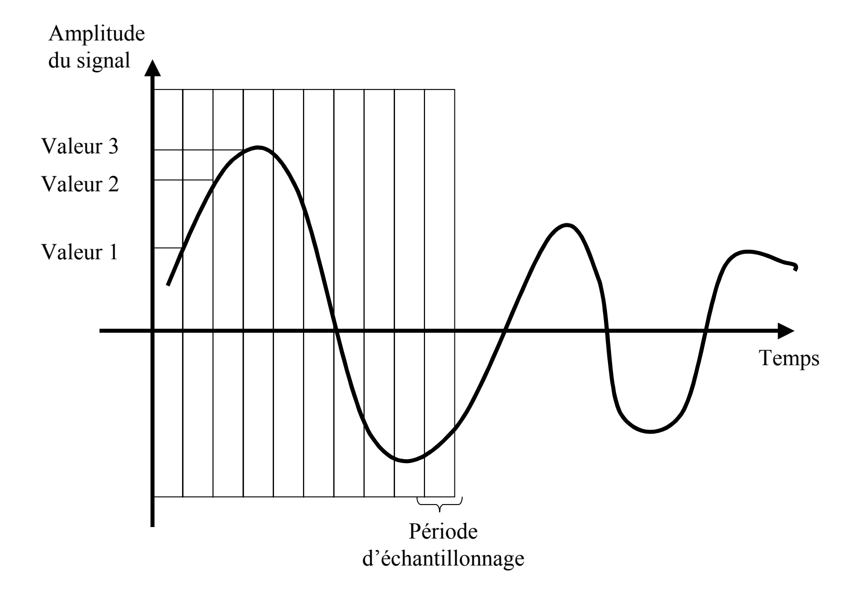


Figure 1.2. *Echantillonnage d'un son*

Le nombre de périodes d'échantillonnage par seconde s'appelle la fréquence d'échantillonnage. Elle doit être suffisamment grande, pour relever suffisamment de valeurs et préserver la forme du signal.

La loi mathématique de Shannon et Nyquist établit que la fréquence d'échantillonnage doit être supérieure ou égale au double de la fréquence maximale contenue dans ce signal. Notre

oreille percevant les sons jusqu'à 20 000 Hz environ, il faut donc une fréquence d'échantillonnage supérieure à 40 000 Hz pour obtenir une qualité satisfaisante. Un certain nombre de fréquences d'échantillonnage sont normalisées :

– 32 kHz : pour la radio FM en numérique (bande-passante limitée à 15 kHz),

– 44.1 kHz : pour l'audio professionnelle et les compact discs.

1.4.5.1.2. Le nombre de bits d'un échantillon

Il détermine le nombre de valeurs que l'échantillon peut prendre. Plus le nombre de bits est important, plus il sera possible de définir des valeurs distinctes de l'amplitude du signal.

Ainsi, avec un codage sur 8 bits, on a 2^8 possibilités, c'est-à-dire 256 valeurs possibles. Avec un codage sur 16 bits, on a 2^{16} possibilités, c'est-à-dire 65 536 valeurs possibles La seconde représentation donnera une qualité de son bien meilleure, mais aussi un besoin en mémoire beaucoup plus important.

1.4.5.1.3. Le nombre de voies

Une seule voie correspond à du mono, deux voies à de la stéréo et quatre voies à de la quadriphonie. La stéréophonie nécessite deux canaux sur lesquels on enregistre individuellement un son qui sera fourni au haut-parleur de gauche, ainsi qu'un son qui sera diffusé sur celui de droite.

Compte tenu des trois éléments expliqués précédemment, la taille en octets d'une séquence sonore se calcule de la façon suivante :

Taille (octets) = Taux d'échantillonnage x Nombre de secondes x Nombre d'octets par échantillon x Nombre de voies

Par exemple, un morceau de 3 mn (180 s) échantillonné à 44,1 kHz sur 16 bits (2 octets) et en stéréo occupera une place de : 44 100 x 180 x 2 x 2 = 31,8 Mo. Il s'agit de la place occupée par cette séquence sonore sans compression.

1.4.5.2. *La compression du son*

Comme indiqué au paragraphe précédent, la place occupée par une séquence sonore n'est pas négligeable. Cela a conduit à chercher des techniques de compression. Le format le plus connu est le MP3 (MPEG Audio layer 3). Ce format permet de compresser à un taux de 1:12 les formats audio habituels (WAV ou CD audio).

La compression consiste à retirer des données audio les fréquences inaudibles pour l'auditeur moyen dans des conditions habituelles d'écoute. Il y a donc perte d'information, mais l'on ne s'en aperçoit pas.

L'essor du multimédia et d'internet ont répandu l'utilisation du son sous forme numérique. Des sites comme Fnac.com ou Amazon.com permettent par exemple d'écouter des séquences de CD *via* internet.

Notons que les sites « peer to peer » permettant d'échanger gratuitement des fichiers de musique numérique portent gravement atteinte aux auteurs. Ce type de téléchargement est devenu illégal et l'on s'oriente peu à peu vers des sites payants, comme par exemple i-tunes ou Fnac.com.

1.5. La structuration des données

C'est un fait, le volume de données stockées et échangées ne fait que croître. Cette prolifération conduit à organiser les données de façon structurée dans des bases de données pour pouvoir les stocker et y accéder facilement.

1.5.1. *Evolution de la technologie et du stockage*

La technologie a considérablement évolué depuis la création du premier ordinateur. Cette évolution s'est faite (et continue de se faire) dans le sens d'une :

– miniaturisation des serveurs,
– augmentation de la puissance des processeurs,
– augmentation des capacités de stockage,
– diminution des coûts.

Le tableau 1.2 résume cette évolution au cours des 25 dernières années.

	1980	**1990**	**2000**	**2005**	**Evolution 2005 / 1980**
Processeurs					
Fréquence (Mhz)	1	20	600	4000	X 4000
Mémoires DRAM					
Capacité type (MB)	0,064	4	64	1024	X 16000
Accès (ns)	375	100	60	40	/ 9
Coût ($/MB)	8000	100	1,5	1	/ 8000
Disque					
Capacité type (MB)	1	160	9000	400000	X 400000
Accès (ms)	87	28	8	4	/ 20
Coût ($/MB)	500	8	0,05	0,001	/ 500000

Tableau 1.2. *Evolution de la technologie entre 1980 et 2005*

Les capacités croissantes de stockage conjuguées à la diminution des coûts constituent un facteur évident d'expansion de données stockées. Ainsi, le volume d'informations stockées au sein des entreprises a doublé chaque année entre 2000 et 2003 (source : *EMC*).

Ce phénomène est amplifié avec l'utilisation d'internet : le courrier électronique, les sites web, les fichiers téléchargés, etc. occupent de plus en plus de place.

1.5.2. *Prolifération des sources de données*

Les sources d'information ont augmenté de façon spectaculaire au cours des trente dernières années.

Rappelons qu'à ses débuts, l'informatique était centralisée et les salariés utilisant un ordinateur étaient nettement moins nombreux qu'aujourd'hui. Aussi, le nombre de données gérées était globalement assez faible.

La mise à disposition massive de PC a constitué une étape importante dans la prolifération de données stockées et échangées. Chaque utilisateur d'un PC a dès lors pu créer, stocker, diffuser des données très facilement. Plus autonome, il est aussi devenu plus exigeant, consommant et produisant à tort ou à raison davantage d'informations.

Plus récemment, l'arrivée d'internet et la généralisation des intranet a aussi accru les sources de données, participant là encore à une multiplication de celles-ci. L'explosion du courrier électronique illustre parfaitement cette tendance : en 2004, près de huit mille milliards d'e-mails professionnels ont été échangés, nécessitant une capacité de stockage globale de cent cinquante mille To (Tera octets) !

La technologie sans fil va vraisemblablement participer à ce mouvement dans les années à venir : sans la contrainte d'un poste fixe, il est désormais possible de consulter, modifier, échanger des informations n'importe où et n'importe quand.

C'est donc l'évolution de la technologie permettant une utilisation sans contrainte qui explique la prolifération des données.

1.5.3. *Les bases de données*

Comme le montre le paragraphe précédent, le nombre de données manipulées ne fait qu'augmenter. La question qui se pose est donc de savoir comment stocker ces montagnes d'informations et surtout y accéder le plus facilement possible ? La réponse réside dans l'utilisation de bases de données. De façon très simplifiée, une base de données est un conteneur structuré de données.

Pratiquement, les bases de données sont intégrées dans un ensemble de logiciels permettant de décrire, de créer, de modifier, de supprimer et de consulter les données. La base de données et ses logiciels associés constituent le Système de gestion de base de données (SGBD). En termes d'architecture, le modèle relationnel s'est nettement imposé et l'on parle couramment aujourd'hui de SGBDR (Système de gestion de base de données relationnelles).

1.5.3.1. *Structuration de l'information*

Le nombre important de données manipulées nécessite de les grouper logiquement dans des entités homogènes. Par exemple, les données clients sont groupées dans une même entité « Client ». Les données article sont groupées dans une autre entité « Article ».

Cette structuration résulte d'un travail de modélisation des données au cours duquel il convient de :

– identifier des données que l'on souhaite gérer ;

– définir l'utilisation de ces données (saisie, consultation, modification…) ;

– grouper les différentes données au sein d'entités logiques ;

– définir les liens entre ces entités : par exemple, spécifier qu'un article peut être vendu à plusieurs clients (et inversement, qu'un client peut commander plusieurs articles).

Beaucoup de méthodes de modélisation existent et reposent pour la plupart sur le modèle conceptuel de données (MCD) appelé aussi modèle entité-relation.

Nous présentons ci-après (figure 1.3) un exemple de modèle conceptuel des données. Cette représentation met en évidence les objets et les relations gérés dans le processus d'une commande d'achat passée à un fournisseur. L'exemple présenté est simplifié et a pour seul objectif d'expliquer le principe du modèle entité-relation.

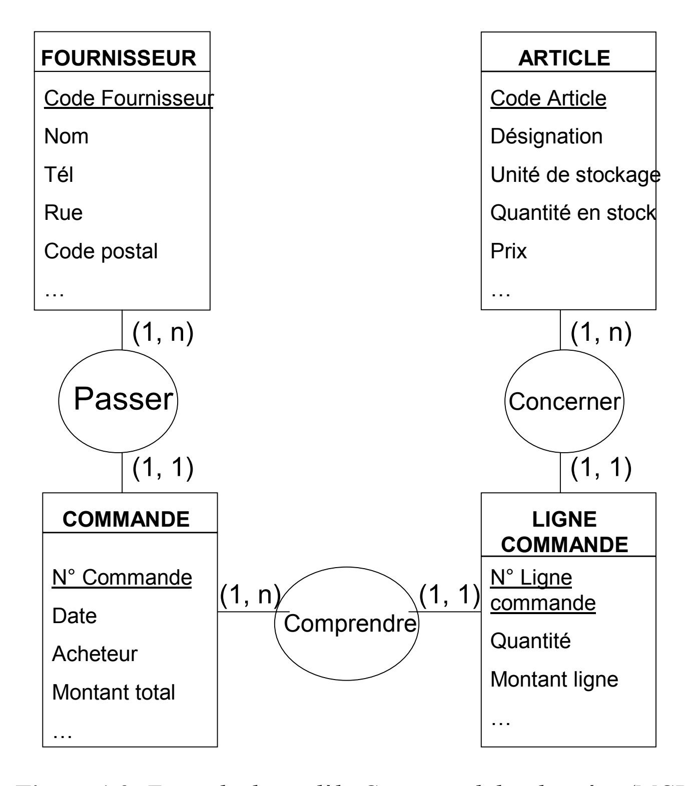


Figure 1.3. *Exemple de modèle Conceptuel des données (MCD)*

Cet exemple met en évidence les règles de gestion suivantes :

– quatre entités interviennent dans la représentation : le fournisseur, la commande d'achat, la ligne de commande et

l'article ; pour chacune des entités, un certain nombre de données sont listées ; il s'agit des attributs de l'entité ; le premier attribut souligné correspond à l'identifiant dont le but est de différencier chaque fournisseur, chaque article, etc. ;

– trois relations « passer », « comprendre » et « concerner » traduisent les liens logiques qu'il y a entre les entités ; les liens peuvent aussi être porteurs d'information, donc d'attributs ;

– les cardinalités, par exemple (1, *n*) reflètent les règles de gestion :

- une commande d'achats donnée est passée à un et un seul fournisseur ; réciproquement, un fournisseur donné peut faire l'objet de plusieurs commande d'achats ;

- une commande d'achats peut comprendre 1 à *n* lignes de commandes ; réciproquement une ligne de commande donnée ne peut faire partie que d'une et une seule commande ;

- une ligne de commande concerne un et un seul article ; réciproquement un article donné peut faire l'objet de plusieurs de commandes.

Une fois modélisées, les données et les relations ainsi structurées se traduisent par des tables dans le SGBDR. Ces tables peuvent alors accueillir des données selon la structuration et les règles de gestion établies lors de la modélisation.

De nombreux ouvrages traitent de façon détaillée des méthodes de conception et de modélisation. Ces méthodes sont particulièrement adaptées au développement d'applications spécifiques, pour lesquelles il y a besoin de construire la base de données. Elles reposent pratiquement toutes sur le modèle entité-relation décrit précédemment. La méthode Merise fut pionnière en la matière dans les années 1970 et, depuis, bien d'autres méthodes ont vu le jour.

En revanche, les progiciels intègrent en standard un modèle de données et l'effort de conception est ainsi économisé. Il est cependant nécessaire de comprendre le modèle de données et les règles de gestion sous-jacentes du progiciel, pour pouvoir

étudier l'adéquation de ce modèle avec les besoins de l'entreprise.

1.5.3.2. *Stockage centralisé et non-redondance des informations*

Dans les SGBDR une donnée n'est stockée qu'à un seul endroit. Même si physiquement, les données peuvent être réparties sur des installations matérielles différentes, l'administration des données est centralisée. L'utilisateur accède à l'information comme si celle-ci n'était stockée qu'à un seul endroit.

Ainsi, le SGBDR facilite la non-redondance des données : des utilisateurs différents peuvent partager les mêmes données.

1.5.3.3. *Indépendance des données et des traitements*

Les données sont totalement indépendantes des programmes qui les utilisent. La base de données est stable et différents programmes peuvent tout à fait exploiter les mêmes données. Inversement, on peut imaginer de remettre en cause les données sans pour autant modifier les programmes.

Cependant, si des évolutions de la base de données sont toujours possibles, comme par exemple l'ajout d'un nouvel attribut, il faut être conscient que les enregistrements créés avant l'évolution doivent être mis à jour si l'on veut qu'ils bénéficient de l'évolution de structure. Prenons l'exemple d'une base de données clients, comprenant à un instant *t* 3 000 enregistrements.

Supposons que l'on souhaite ajouter l'attribut « Stratégique » pouvant prendre les valeurs « O » (stratégique) ou blanc (non stratégique). Si l'on ne fait rien, les 3 000 clients présents dans la base de données seront considérés comme non stratégiques, ce qui ne correspond évidemment pas à la réalité. Dans cet exemple, l'ajout d'un nouvel attribut doit donc être accompagné d'un programme de reprise permettant de mettre à jour les enregistrements existants de la base de données, de façon à

déclarer les clients stratégiques de la base existante. En outre, il y a besoin de modifier l'écran de saisie des clients pour faire apparaître la nouvelle zone et permettre sa saisie lors des futures créations de clients. Enfin, les requêtes et éditions doivent aussi évoluer si l'on souhaite y faire apparaître la nouvelle donnée.

Il y a donc bien indépendance des données et des traitements, mais toute évolution de la structure de la base de données engendre néanmoins, d'une part, un travail de mise à jour des enregistrements existants, d'autre part, une adaptation des écrans de saisie, requêtes, éditions.

1.5.3.4. *Sécurité des données*

De nombreux outils livrés avec les SGBDR garantissent la sécurité des données. Si la sécurité des données est souvent perçue comme synonyme de confidentialité, d'autres caractéristiques sont à prendre en compte. La sécurité recouvre en particulier :

– la confidentialité : l'accès aux données peut être filtré en fonction des utilisateurs ; par exemple, les données comptables d'un client pourront être consultées par l'ensemble des utilisateurs de l'entreprise, mais seul le service comptabilité clients pourra créer, modifier ou supprimer ces informations ;

– la fiabilité : différents mécanismes de sauvegarde et de reprise permettent de restaurer les données dans l'état précédant un incident ; autrement dit, il est toujours possible de repartir d'une « photographie » de la base de données avant incident ; bien entendu, les données saisies dans l'intervalle de temps compris entre cette photographie et l'incident doivent être ressaisies, ce qui pose souvent problème ; comment se souvenir de tout ce qui a été effectué entre-temps ? c'est la raison pour laquelle il est vivement conseillé d'imprimer régulièrement les documents marquants de l'activité de l'entreprise pour garder des traces ; par exemple, dans un service administration des ventes, il est recommandé de faire systématiquement des impressions des commandes enregistrées ;

– la traçabilité : en cas de problème important ou d'attaque du système et, avant de restaurer les données dans un état antérieur, il est nécessaire d'analyser les traces, c'est-à-dire la séquence des événements ayant impacté la base de données ; les SGBD permettent ainsi de tracer l'enchaînement des mises à jour, des accès… et de dater précisément quand s'est produit l'incident ; le niveau de détail de ces traces est paramétrable et concerne les aspects système, réseau, et accès aux données élémentaires elles-mêmes ;

– l'intégrité : les SGBD assurent la cohérence des données conformément aux règles de gestion de l'entreprise ; les accès concurrents des utilisateurs en particulier sont gérés de façon à ne pas engendrer d'incohérences.

Notons que la sécurité est une réponse à deux grandes catégories de menaces :

– les attaques accidentelles dues à des défaillances du système (problème matériel, réseau ou autre) ou à des incidents et catastrophes divers (incendie, dégât des eaux, négligence humaine…) ;

– les attaques malveillantes qui visent à détruire volontairement tout ou partie du système d'information ; les attaques les plus connues sont les virus qui détruisent des données et se propagent *via* les réseaux informatiques ; cette seconde catégorie tend à croître depuis quelques années.

1.5.4. *Le langage XML (*eXtensible Markup Language*)*

Le langage XML (langage à balise extensible) a fait franchir une nouvelle étape à Internet. Il a été mis au point en 1996 et reconnu comme standard par le W3C (World Wide Web Consortium) dès 1998.

Au début d'internet, le World Wide Web reposait uniquement sur le langage HTML (*Hypertext Markup Language*). Ce langage, encore largement utilisé aujourd'hui, est basé sur du

texte et des « hyperliens » permettant de naviguer d’une page à une autre. L’inconvénient de HTML est qu’il n’y a que du texte. L’ordinateur peut donc uniquement lire, mémoriser et imprimer les pages HTML.

Avec le langage XML au contraire, *les données sont structurées*. Et c’est bien là tout l’intérêt de XML, car les informations peuvent être traitées et stockées dans des bases de données. Aussi, la capacité de rechercher et de transmettre des informations devient beaucoup plus efficace et facilite le commerce électronique. De façon simplifiée, on peut assimiler XML au langage de l’EDI (*Electronic Data Interchange*), avec tous les avantages du web.

Sans entrer dans le détail, un document XML est composé de trois parties :

– le prologue (obligatoire) qui permet d’indiquer la version de la norme XML utilisée pour créer le document ainsi que le jeu de caractères utilisé dans le document ;

– la DTD (*Document Type Definition*, optionnel) qui décrit la structure hiérarchique du document ;

– le contenu physique des données et balises (obligatoire).

Les applications de XML sont nombreuses :

– publication de documents : la présentation des données d’un document peut facilement être adaptée en fonction du support (PC, téléphone mobile…) ;

– échange de données,

– serveurs web XML,

– bases de données XML,

– interconnexion de progiciels,

– portails d’entreprises,

– application Business to Business.

Le succès de XML s'explique par ses principaux atouts :

– il peut être facilement interprété par les individus ;

– il véhicule uniquement des données et est indépendant de la plate-forme, du protocole de communication, du langage et du système d'exploitation ;

– il est utilisé de façon native par les grandes bases de données et peut avoir des applications dans tous les domaines.

1.6. Conclusion

Ce premier chapitre a introduit le rôle essentiel de l'information au sein des organisations. L'information est devenue une matière première indispensable à la vie et au développement des entreprises. Pratiquement tous les métiers sont concernés. La qualité et la maîtrise des données sont donc un enjeu majeur pour rester dans la compétition. Privée d'information, aucune entreprise ne peut survivre.

Les données multimédias sont encore peu utilisées dans le domaine de la gestion d'entreprise et l'on peut s'attendre à une augmentation de la place occupée par ce type de données (images, vidéo, son) au cours des prochaines années, avec des disparités selon le secteur d'activité. Aussi, la diffusion de l'information sous ces différentes formes est désormais accélérée par le développement d'internet.

L'explosion des sources de données, favorisée par les progrès technologiques, a conduit à organiser et structurer l'information au sein de bases de données pour y accéder plus facilement. Les bases de données relationnelles sont aujourd'hui implantées partout au sein des organisations. Les principaux fournisseurs de SGBDR sont Oracle, IBM et Microsoft. Avec le développement d'internet, le langage XML est devenu un standard de structuration et d'échanges de documents.

CHAPITRE 2

Tour d'horizon des principales tendances informatiques

Ce chapitre présente les grandes tendances en matière d'applications informatiques. Il permet de mieux comprendre l'enjeu des données pour chacun de ces domaines. Très clairement, ce chapitre n'est pas exhaustif car il ne couvre pas l'immense champ des multiples applications informatiques que l'on peut rencontrer. Il se concentre simplement sur les grands courants rencontrés le plus fréquemment au sein des organisations.

2.1. Les progiciels de gestion intégrés (ERP)

Les PGI (Progiciels de gestion intégrés en français) appelés couramment ERP (*Enterprise Resource Planning*), se sont largement implantés au sein des grandes entreprises au cours des quinze dernières années. Le terme « Progiciel » résulte d'une concaténation des termes Produit et Logiciel. Un progiciel intégré est donc un produit logiciel conçu et vendu par un éditeur, couvrant plusieurs domaines fonctionnels de façon intégrée.

Les différentes fonctions de l'entreprise (comptabilité, contrôle de gestion, vente, production, maintenance, achats, stocks, ressources humaines…) sont donc prises en charge par l'ERP grâce à des *modules* fonctionnels. L'*intégration* est assurée par une base de données relationnelle partagée par les différents modules.

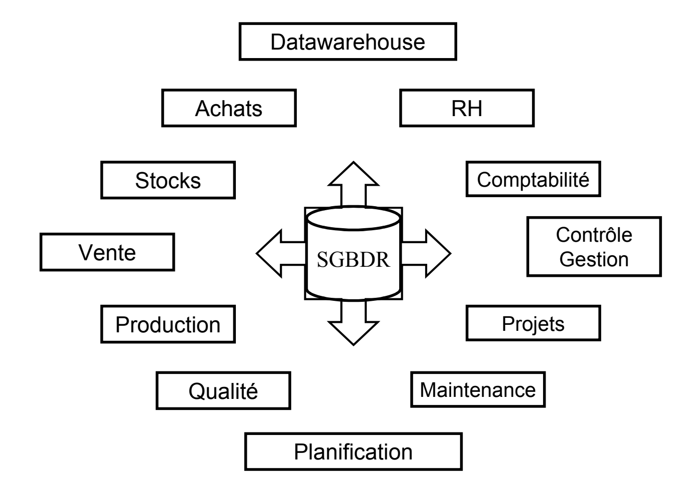


Figure 2.1. *Domaines couverts par un ERP*

2.1.1. *Les raisons de l'essor des ERP*

Le succès des ERP s'explique par plusieurs facteurs :

– recentrage des entreprises sur leur métier d'origine : dans un environnement économique difficile, les entreprises ont tendance à se recentrer sur leur métier d'origine ; toutes les directions et services sont concernés par ce mouvement de recentrage ; les directions informatiques ont largement été impliquées dans cette démarche et les développements d'applications spécifiques souvent longs, coûteux et incertains ont laissé peu à peu place à l'implémentation de solutions packagées de type progiciel métier ou ERP ;

– optimisation des processus : dans chaque domaine fonctionnel, l'ERP s'appuie sur les meilleures pratiques de

gestion ; aussi, la mise en place d'un ERP a l'avantage de réduire voire supprimer les coûts liés au développement du logiciel, puisque le progiciel est déjà conçu et développé ; l'entreprise peut alors se consacrer au paramétrage du progiciel (pour que celui-ci « colle » à son organisation et à sa stratégie), aux procédures à mettre en place, à la formation des utilisateurs, à la conduite du changement, etc. ; la mise en place d'un ERP est souvent l'occasion de repenser et améliorer les processus de l'entreprise ;

– intégration et cohérence transversale des données de l'entreprise : c'est incontestablement un point fort des ERP ; s'appuyant sur un SGBD relationnel, les données sont partagées par toutes les fonctions et ne sont saisies qu'une seule fois à la source ; les échanges entre différents services ou départements d'une entreprise sont fluidifiés ; l'ERP joue donc un rôle moteur dans le décloisonnement des structures de l'entreprise, qui devient ainsi plus agile ;

– maîtrise des coûts informatiques : les coûts liés à la technique informatique sont diminués au profit d'un investissement dans l'optimisation des processus ; la mise en place d'un ERP se fait généralement avec l'aide de prestataires extérieurs ; cette sous-traitance permet clairement une meilleure maîtrise des coûts, tout en s'appuyant sur de solides compétences produit et métier.

Les raisons de ce succès ne doivent cependant pas masquer les obstacles parfois rencontrés lors de la mise en place d'un ERP. L'impact sur l'organisation et les processus nécessite généralement un effort d'adaptation non négligeable pour l'entreprise.

2.1.2. *Les facteurs-clés de succès d'un projet ERP*

La réussite d'un projet ERP repose globalement sur :

– l'implication du top management de l'entreprise : l'ERP touche de près à l'organisation, aux structures, aux processus ; la

mobilisation et l’adhésion des collaborateurs de l’entreprise est une condition *sine qua none* à la réussite du projet ; il est donc essentiel que le top management soit moteur tout au long du projet et que la ligne directrice soit clairement tracée par les dirigeants ; la motivation et l’appropriation de l’ERP sont à la hauteur de l’implication des managers ;

– une équipe projet de qualité : les équipes projet doivent être de grande qualité car une forte implication est requise ; la maîtrise du produit, la compréhension du métier et des processus, le respect des délais, des facultés d’adaptation et de communication sont nécessaires ; aussi, il convient de ne pas négliger la dimension technique (architecture matérielle et système, réseaux…) avec la mise en place de compétences spécifiques dans ce domaine ; enfin, l’équipe de management du projet joue un rôle-clé dans la réussite du projet (planification, budget, suivi, arbitrages, gestion des compétences…) ; cet accompagnement dans l’expertise produit, métier, technique et management est généralement réalisé par ou avec l’aide de consultants expérimentés dans ce type de produit et de projet.

Le marché des ERP est largement dominé par SAP. De nombreux autres éditeurs se partagent le reste du marché : Oracle, Peoplesoft (racheté depuis peu par Oracle), Intentia, JD Edwards, Adonix… et bien d’autres.

Les grandes entreprises ont pour la plupart adopté l’un des grands ERP du marché et l’on assiste aujourd’hui à des offres ciblées de la part des éditeurs en direction des PME.

2.2. Le commerce électronique (*e-business*)

Avec l’arrivée d’internet et des technologies associées, un certain nombre d’entreprises ont développé des applications permettant d’effectuer des ventes par ce biais : c’est le commerce électronique.

2.2.1. *L'expansion du commerce électronique*

Malgré l'éclatement de la bulle spéculative en mars 2000, le commerce électronique connaît depuis plusieurs années un développement spectaculaire. Ainsi, en 2004 on a enregistré en France une progression de 53 %. Les Français ont dépensé six milliards d'euros en 2004 dans les achats en ligne (hors commerce inter-entreprises et services financiers) contre 3,6 milliards d'euros pour 2003, selon les calculs de la Fédération des entreprises de vente à distance (Fevad).

Cette tendance se confirme chez les grands noms français du secteur (+ 100 % pour la Redoute en 2004, + 71 % pour Voyages-Sncf et + 52 % pour la Fnac). Le commerce destiné aux particuliers est le plus dynamique. Les échanges B to C (c'est-à-dire s'adressant au consommateur) affichent une progression de 58 %, les ventes interentreprises c'est-à-dire les échanges B to B (business to business) ont, quant à elles, progressé de 45 %. Le tableau 2.1 présente le « Top 10 » de l'e-commerce au mois de janvier 2005 :

Chiffres de janvier 2005	
Sites	**Nombre de visiteurs uniques (en millions)**
Ebay.fr	4,687
Kelkoo.com	4,021
Fnac.com	3,908
Voyages-Sncf.com	3,835
Cdiscount.com	3,352
Amazon.fr	2,593
Priceminister.com	2,442
Laredoute.fr	2,315
Alapage.fr	2,115
Rueducommerce.fr	1,832

Source : Challenges / Médiamétrie / NetRatings

Tableau 2.1. *Top 10 du commerce électronique en janvier 2005*

Le commerce électronique s'est d'abord développé dans le secteur de la vente d'ordinateurs, des logiciels, des CD, des livres ou encore des voyages. Ainsi, Amazon, eBay, la Fnac, Dell, Air France, SNCF… sont quelques exemples de réussites de sites de commerce électronique. Et il y en a bien d'autres !

On constate aujourd'hui que d'autres secteurs d'activité développent aussi leurs ventes par ce biais, mais dans des proportions encore relativement faibles. Un certain nombre d'experts estiment cependant que la plupart des secteurs d'activité se tourneront tôt ou tard vers le commerce électronique.

2.2.2. *Les raisons du succès*

Le commerce électronique présente un certain nombre d'avantages évidents, pour le vendeur, comme pour l'acheteur, parmi lesquels :

– pas de contrainte liée au temps : les clients peuvent acheter à tout moment, sans la contrainte de rencontrer un commercial ou un vendeur pendant les heures de bureau ;

– pas de contrainte géographique : une simple connexion internet permet aux clients d'acheter les produits ;

– meilleure interactivité client- fournisseur : les achats en ligne du client sont mémorisés ; les « clics » effectués par l'acheteur constituent une source précieuse d'information et permettent au vendeur de mieux connaître ses clients et de personnaliser ainsi son offre ; cette personnalisation améliore la satisfaction client et facilite la *fidélisation* ;

– dans certains cas, outre la facilité d'effectuer des achats, internet offre des services supplémentaires par rapport aux achats classiques ; ainsi les supermarchés en ligne évitent au client de se déplacer.

Sur le plan informatique, les sites de commerce électronique comprennent :

– une connexion avec les bases de données de l'entreprise (système back-office) ; c'est un point essentiel, car les données communiquées sur internet doivent être fiables, sous peine de voir les clients disparaître à jamais ; par exemple, les données prix, disponibilité des stocks, date de livraison doivent être correctes, ce qui suppose une connexion temps réel avec le back-office ;

– un système de paiement sécurisé : incontournable pour garantir la sécurité et la confidentialité du paiement ; si les questions de sécurité constituaient un obstacle majeur au commerce électronique il y a quelques années, les clients acceptent à présent les niveaux de sécurité actuels ; d'ailleurs les experts de ce domaine s'accordent pour reconnaître qu'il y a bien plus de risque à communiquer son numéro de carte bleue par téléphone ou à jeter une facturette après utilisation de sa carte bancaire.

Notons enfin que le succès d'un site de commerce électronique repose en grande partie sur la qualité de l'interface utilisateur. Le site doit être séduisant et convivial, comporter une présentation attrayante avec une disposition de textes et d'images très claire et, bien entendu, offrir des temps de réponse excellents.

2.3. Les systèmes d'aide au pilotage (*Business Intelligence*)

2.3.1. *La définition des indicateurs*

Les systèmes d'aide au pilotage, que l'on regroupe depuis quelques années sous l'appellation de *Business Intelligence* ont pour objectif de synthétiser l'information pour répondre aux besoins des managers. Ces informations synthétiques appelées indicateurs permettent aux managers d'avoir une connaissance

ciblée d'un certain nombre de composantes de leur activité, les aidant ainsi à mieux adapter leur stratégie et leurs actions.

La définition des indicateurs est une tâche essentielle dans la mise en place d'un tel système. Ceux-ci doivent en effet être déterminés de façon à restituer des informations pertinentes et utiles pour la conduite du business. Un indicateur pourra comporter différents axes d'analyse qu'il convient de définir lors de la conception du système. Par exemple, l'indicateur chiffres d'affaires pourra être analysé selon les axes géographique, produit, client, temps… que l'on pourra combiner. La combinaison des axes permet ainsi de faire ressortir un certain nombre de caractéristiques de l'activité analysée et de mieux agir.

Les indicateurs sont généralement calculés à partir des données des systèmes d'information opérationnels et l'on comprend aisément qu'il est indispensable que ces derniers fournissent des informations de qualité.

2.3.2. *Collecte, stockage et restitution des informations*

En pratique, les données des systèmes de pilotage sont stockées dans des bases de données dédiées appelées *Datawarehouse* (entrepôt de données), car les systèmes opérationnels délivrent une grande quantité de données détaillées, dont beaucoup sont inutiles aux systèmes de pilotage. Aussi, les données stockées dans l'entrepôt de données doivent être homogènes, bien que provenant généralement de différentes sources. Ce travail de collecte et d'homogénéité est l'une des fonctions-clés des systèmes de pilotage. Enfin, à l'inverse des systèmes opérationnels, les systèmes d'aide au pilotage conservent un historique sur plusieurs années, pour pouvoir analyser et comparer les informations sur une période relativement longue. Généralement, le Datawarehouse repose sur un modèle de données dit dimensionnel, adapté à la définition de plusieurs axes d'analyse (le modèle relationnel

peut aussi être utilisé, bien qu'il soit davantage adapté aux applications opérationnelles).

La collecte et la transformation des données est généralement assurée par des outils spécialisés appelés ETL (*Extract Transform and Load*). Ces outils permettent d'extraire des informations provenant de systèmes hétérogènes pour les charger dans la base de données décisionnelle après adaptation (voir section 8.4).

La restitution des informations à l'utilisateur se fait soit *via* des états prédéfinis qui sont exécutés et diffusés périodiquement, soit par des requêtes qui peuvent être lancées à la demande. Les requêtes comprennent généralement :

– un écran permettant de définir des critères de sélection de façon à indiquer le périmètre de l'analyse (par exemple, telle zone géographique, tel produit ou tel groupe de produits) ;

– un écran ou une fenêtre permettant de définir les données à restituer.

2.3.3. *Principaux domaines d'application*

Les outils actuels offrent une bonne ergonomie (interface graphique, présentation *look and feel*…) permettant de naviguer facilement au sein des axes d'analyse. Ainsi, le chiffre d'affaires d'une division peut facilement être analysé selon l'axe géographique et l'on peut en quelques clics descendre au niveau d'un pays, d'une région, d'une agence commerciale. Cette facilité de manipulation explique en grande partie le succès que remportent actuellement les systèmes de business intelligence.

Les indicateurs de type financiers sont de loin les plus répandus, suivis des indicateurs opérationnels, puis des indicateurs de suivi clients. Il y a en revanche une quasi-absence d'indicateurs sur la concurrence, la conjoncture et les tendances macroéconomiques. Pourtant, ces indicateurs relatifs à

l'environnement constituent un avantage concurrentiel incomparable. Cette lacune devrait donner lieu à des projets dans ce domaine dans les années à venir.

Soulignons enfin que dans beaucoup d'entreprises, on estime qu'il y a trop d'indicateurs et que peu d'entre eux sont réellement utiles. Ce constat illustre la difficulté à définir les indicateurs vraiment pertinents. Aussi, il est recommandé de s'appuyer sur des méthodologies solides lors de la mise en place d'outils de pilotage. Actuellement le modèle « Balanced Scorecard » s'est affirmé dans de nombreuses entreprises américaines. D'autres modèles tels que « Total Quality Management », « Six Sigma » ou encore la méthode ABC sont également très prisés.

2.4. Les systèmes de gestion de la relation client (CRM)

L'évolution récente du Marketing vers une gestion de la relation client (GRC) ou en anglais CRM (*Customer Relationship Management*) s'explique, d'une part, par la volonté de mieux maîtriser les interactions avec les clients en vue d'accroître le volume de ventes totales, d'autre part, par la recherche d'une diminution des coûts relatifs à cette gestion.

2.4.1. *La connaissance des clients*

La connaissance des clients est fondamentale dans le CRM. Elle nécessite la mise en place d'un système d'acquisition et de stockage d'informations adapté au secteur d'activité. Le système d'acquisition des données clients doit en effet être pensé en fonction de l'exploitation que l'on souhaite en faire et des possibilités d'acquisition des données propres au secteur d'activité.

Dans le cas d'une banque par exemple, les retraits par carte ou les opérations au guichet constituent une source

d'informations essentielles au banquier. En combinant ces données avec le profil sociodémographique du client (adresse, âge, profession, statut matrimonial…) obtenu lors d'une demande de prêt par exemple, on peut définir les produits les plus adaptés à proposer en priorité. Cela permet aussi d'évaluer le risque financier du client, de connaître ce que rapporte un client sur une période donnée, etc.

La grande distribution en revanche ne connaît pas ses clients car le passage en caisse est anonyme. La mise en place de cartes de fidélité ou de paiement permet de combler cette lacune, en associant le détail des achats lors d'un passage en caisse à une carte donnée, donc à un client. Aussi les informations sociodémographiques obtenues lors de l'adhésion permettent au distributeur de mieux connaître le client et de pouvoir adapter sa communication (offres ciblées, promotions…).

Sur le plan informatique, toutes ces données sont stockées dans un entrepôt de données (Datwarehouse) intégré dans le paysage applicatif de l'entreprise. La fréquence de mise à jour de l'entrepôt (journalier, hebdomadaire, mensuel) dépend de l'utilisation qui en est faite.

2.4.2. *L'optimisation de la relation client*

L'optimisation de la relation client est le deuxième volet important du CRM. La centralisation ou l'automatisation des processus réguliers et récurrents permet d'établir une relation à la fois plus forte et moins coûteuse. Les centres d'appels téléphoniques ou le traitement automatisé des courriers sont des exemples classiques de cette optimisation. L'application d'une politique CRM demande à l'entreprise de s'organiser en conséquence en privilégiant une approche globale du client, ce qui n'est pas toujours chose évidente car beaucoup d'entreprises sont encore organisées autour des produits. Il n'est pas rare de voir des équipes marketing d'une même société se trouver en concurrence et proposer à un même client pléthore de produits,

sans arbitrage ni recherche d'offre globale réellement adaptée aux besoins de ce client.

L'optimisation des processus s'appuie généralement sur des logiciels dédiés au CRM (SIEBEL, VANTIVE…) ou sur des modules de progiciels de gestion intégrés (SAP, Peoplesoft…). Il y a également un certain nombre de logiciels de « Datamining » spécialisés dans l'analyse statistique, qui supportent l'analyse prédictive des comportements clients et optimisent la segmentation de clientèle. Notons cependant que l'utilisation des logiciels de Datamining (Enterprise Miner de la société SAS, Intelligent Miner d'IBM…) requiert des compétences pointues en statistiques.

Retenons en conclusion de ce paragraphe que le CRM peut s'avérer être une arme redoutable en termes de compétitivité. Pour être efficace, il y a besoin de repenser l'organisation Marketing en l'orientant vers une approche globale des clients (par opposition à une approche purement produit). Il est nécessaire en outre de s'entourer de solides compétences en informatique mais aussi en statistique et évidemment en marketing.

2.5. Quelques mots sur les réseaux informatiques

Les réseaux informatiques ne jouent pas un rôle déterminant dans la qualité des données. Cependant ils représentent un élément essentiel des systèmes d'information de gestion car ils permettent la diffusion des données d'un point à un autre.

Ainsi, les systèmes de type client-serveur, majoritairement utilisés aujourd'hui, utilisent le réseau pour véhiculer les données du serveur sur lequel se trouvent les base de données vers le poste utilisateur, le client, qui peut prendre en charge une partie du traitement des données. Les réseaux permettent aussi d'échanger des données avec l'extérieur (clients, fournisseurs, sous-traitants, etc.).

Les réseaux sont physiquement constitués de trois parties : les postes de travail (par exemple, les PC des utilisateurs), les serveurs qui hébergent les applications utilisateurs et le système de télécommunication qui assure le transport des données d'un point à un autre (câbles, équipements électroniques).

2.5.1. *Différents types de réseaux*

Les réseaux véhiculent les informations sous forme de messages qui circulent sur un support de transmission (câble, fibre optique, ondes radio). Le découpage des messages en paquets et le transport de ces paquets s'effectue selon une normalisation qui comporte sept couches différentes (le modèle OSI). Les architectures réseaux peuvent être complexes et leur conception et leur mise en œuvre nécessitent des compétences pointues.

Selon l'étendue géographique, on distingue :

– les réseaux locaux, appelés LAN (*Local Area Network*) : ils permettent l'échange de données sur un site d'une entreprise ;

– les réseaux étendus, appelés WAN (*Wide Area Network*) : ils permettent l'échange de données entre des sites éloignés à l'échelle nationale ou mondiale ; par exemple, le siège d'une société situé à Houston au Texas peut échanger des données avec ses filiales réparties aux quatre coins du monde *via* le WAN ; les réseaux étendus nécessitent généralement l'utilisation d'opérateurs de télécommunications (France Télécom ou autres).

On classe aussi les réseaux selon leur usage :

– les réseaux privés : il s'agit typiquement du réseau interne à l'entreprise,
– les réseaux publics : réseau internet par exemple.

Sur le plan technique, les interconnexions peuvent être basées sur le routage, la commutation ou les ondes hertziennes.

2.5.2. *Le routage*

L'acheminement du message s'effectue *via* des routeurs et le chemin est déterminé en fonction du trafic. Les routeurs sont des serveurs dont le rôle est d'orienter les paquets dans le réseau. Le standard le plus connu est Ethernet (il existe aussi l'Ethernet commuté). Le coût de cette technique n'est pas très élevé, mais la qualité de service chute avec l'augmentation du trafic. Aussi, la sécurité n'est pas excellente.

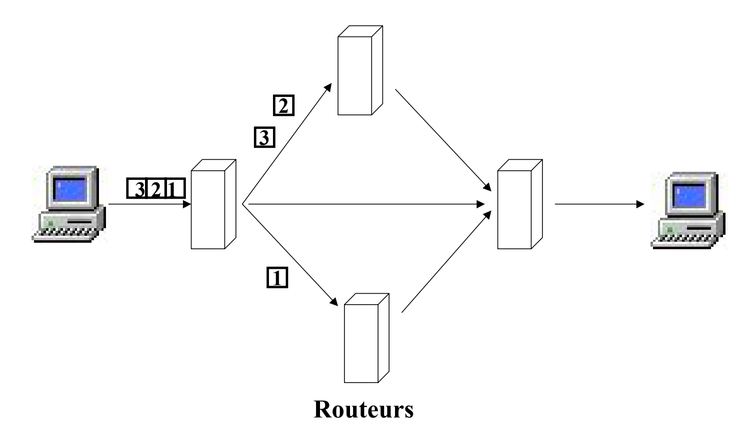


Figure 2.2. *Acheminement des messages par routage*

2.5.3. *La commutation*

L'acheminement du message s'effectue *via* des commutateurs, qui sont des serveurs intelligents. Lors de l'envoi d'un message, il y a détermination d'un chemin optimal et les paquets sont poussés un par un sur ce chemin. Un exemple de réseau commuté est le réseau français Transpac, utilisé en particulier par de nombreuses banques. Cette technique offre une qualité de service nettement meilleure que le routage, ainsi qu'un niveau de sécurité supérieur. Elle est cependant plus coûteuse.

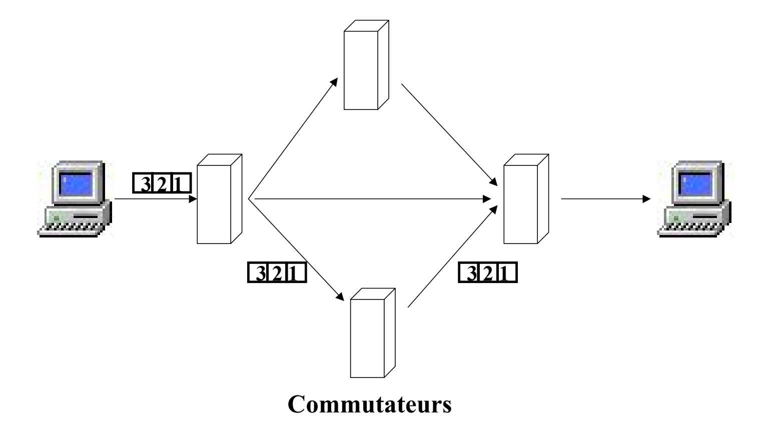


Figure 2.3. *Acheminement des messages par commutation*

Le débit est un facteur important pour la rapidité de la disponibilité des données. Les réseaux commutés offrent un débit beaucoup plus élevé que les réseaux routés. Actuellement, la plupart des réseaux d'entreprise ont un débit de 10 mégabits par seconde.

Bien entendu les besoins en haut débit ne concernent pas toutes les applications. Ainsi, les applications vidéo nécessitent des débits élevés, tandis que l'acheminement de courrier électronique peut se contenter d'un débit relativement faible.

2.5.4. *Les réseaux sans fil*

La diffusion des données sans fil permet de s'affranchir des contraintes liées au câble et apporte une réponse efficace à la mobilité. Les réseaux sans fil connaissent actuellement un développement rapide.

Le principe repose sur l'utilisation des ondes hertziennes et non du câble pour véhiculer les données. Globalement deux normes se détachent du marché :

– Bluetooth : utilisé pour transmettre des données sur de courtes distances avec un débit de l'ordre d'une dizaine de Mbits/sec. Bluetooth se retrouve dans de multiples applications : téléphones mobiles, organiseurs, modems, appareils photo, PC… ;

– IEEE 802.11b alias Wi-Fi (*Wireless Fidelity*) : véritable LAN Ethernet sans fil, le Wi-FI permet de transmettre des données sur plusieurs centaines de mètres avec un débit de plusieurs dizaines de Mbits/sec ; le Wi-Fi peut être utilisé en entreprise comme alternative au réseau local ; aussi, on trouve de plus en plus de « hot spots » sur les lieux publics permettant l'accès à internet depuis un ordinateur portable équipé d'une carte Wi-Fi.

Ces normes évoluent rapidement et des débits de l'ordre de 400 Mbits/sec seront bientôt disponibles (normes 802.11n et UWB – Ultra Wide Band).

2.6. Conclusion

Ce deuxième chapitre illustre quelques grands courants (ERP, commerce électronique, Business Intelligence, CRM) qui se sont imposés au cours des dix dernières années. Chacun de ces domaines repose sur des concepts dont l'objectif est d'améliorer la performance et la compétitivité. Les systèmes d'information sous-jacents ne peuvent cependant répondre à cet objectif que si le contenu, c'est-à-dire l'information qui les nourrit, est de grande qualité.

Quelle que soit la technique (routage, commutation), les réseaux informatiques permettent la diffusion de l'information d'un point à un autre. La mondialisation et les échanges distants de plus en plus répandus s'appuient largement sur les réseaux. La technologie sans fil est prometteuse car elle permet l'accès et l'échange de données sans la contrainte d'un poste fixe ou de câbles.

DEUXIÈME PARTIE

Pourquoi chercher à obtenir des données de qualité ?

CHAPITRE 3

La qualité appliquée aux données

3.1. L'évolution de la qualité

L'histoire de la qualité est liée à l'évolution du rapport producteur/acheteur. La première période de l'industrie est marquée par l'artisanat. La qualité des produits fabriqués était liée à la dextérité de l'artisan, au choix des matières premières et au prix de vente du produit fini, le tout concernant une faible quantité. Comme l'artisan contrôlait régulièrement sa production, la qualité du produit se faisait en temps réel.

Vers la fin du XVIIe siècle, naissent les premières grandes fabriques, fruit du regroupement d'artisans et du progrès des techniques. Comme l'offre reste très inférieure à la demande, tout se vend même ce qui est de mauvaise qualité. Durant cette période, la qualité reste encore liée au prix que l'on paie lors de l'achat : plus un produit est cher, meilleure est sa qualité.

Au début du XXe siècle, l'évolution des techniques entraîne le perfectionnement des machines auxquelles on a de plus en plus recours. Cela se traduit par une forte augmentation des quantités produites. Aussi apparaissent les chaînes de production industrielles qui emploient du personnel peu qualifié et généralement mal payé. Les tâches se divisent en éléments

simples et répétitifs : c'est le taylorisme. La qualité n'est alors liée qu'à un *contrôle* en bout de chaîne.

Dans les années 1950, on assiste à un accroissement spectaculaire des besoins. La qualité des produits devient une composante importante et l'on fait appel à des méthodes de contrôle statistique. Mais la complexité des schémas de production fait grimper le coût des contrôles et les méthodes statistiques atteignent rapidement leurs limites. Par ailleurs, la rentabilité devenant de plus en plus pressante, on passe alors à la qualité par prévention : *l'assurance qualité*.

A partir des années 1960, la qualité est une préoccupation qui se répand dans toute l'entreprise. Elle est désormais prise en compte très en amont du processus de fabrication, dès l'élaboration du cahier des charges. Et comme la concurrence devient de plus en plus rude, on commence à s'orienter vers les besoins des clients.

Dans les années 1970, les produits japonais inondent les marchés avec une qualité meilleure pour un prix moindre. Les clients deviennent de plus en plus exigeants. Et comme l'offre est supérieure à la demande, il faut se battre sur tous les plans pour vendre. La qualité devient à la fois un facteur de compétitivité et un argument commercial.

Aujourd'hui, la qualité est devenue globale et une priorité des entreprises. Elle concerne toutes les activités, tous les métiers, toutes les personnes, tous les instants.

3.2. La qualité appliquée aux données

Dans l'industrie, la qualité d'un produit est correctement appréciée en comparant ses effets aux attentes des clients. Un produit ou service répondant parfaitement aux besoins des clients est donc réputé de grande qualité.

Il en est de même pour les données : une donnée est de qualité si elle répond parfaitement aux besoins des utilisateurs de cette donnée.

Remarquons d'abord que la qualité des données est dépendante de l'utilisation que l'on en fait, c'est-à-dire des utilisateurs. D'une certaine façon, on peut dire qu'une donnée est de qualité si l'ensemble des utilisateurs de cette donnée la jugent ainsi. La compréhension des besoins des utilisateurs est donc une condition nécessaire à la définition et à l'obtention de données de qualité.

Aussi, comme déjà évoqué dans ce livre, une donnée est parfois partagée par des groupes d'utilisateurs différents. La donnée peut être qualifiée de très bonne qualité par un groupe d'utilisateurs et décrite comme de mauvaise qualité par un autre groupe d'utilisateurs. C'est une des difficultés des données partagées.

Enfin, pour un certain nombre d'utilisateurs, une donnée « juste » est synonyme de qualité. La justesse est une condition nécessaire à la qualité des données, mais elle n'est pas suffisante. Car d'autres caractéristiques telles que la compréhension de l'information, le niveau de détail adéquat, son accessibilité, etc. contribuent à la qualité de la donnée. Globalement, la qualité des données se juge selon plusieurs axes dont les plus importants sont certainement :

– qualité du contenu,
– accessibilité,
– flexibilité,
– sécurité.

Ces différentes dimensions sont en grande partie déterminées lors de la conception du modèle de données (voir MCD, paragraphe 1.5.3.1). Le travail de définition du contenu des attributs des objets de données doit donc être effectué avec soin, y compris si un progiciel est utilisé.

3.2.1. *Qualité du contenu*

La qualité du contenu peut être évaluée en fonction des caractéristiques décrites ci-après.

3.2.1.1. *Justesse de l'information*

C'est vraisemblablement la caractéristique la plus importante. Une donnée est juste si elle est en phase avec la réalité. Prenons l'exemple d'une commande client. La quantité 1 000 (pièces) saisie dans le système est juste si le client a effectivement passé une commande de 1 000 pièces. Si le client a passé une commande de 100 pièces et que l'on a saisi 1 000 pièces, la donnée n'est évidemment pas juste. La justesse s'applique à toutes les données et non uniquement aux données numériques. Par exemple, la base de données du personnel comprend l'attribut « sexe ». Si Henri Dupont est enregistré dans cette base avec le sexe « F » (femme), alors cette donnée est incorrecte.

3.2.1.2. *Adéquation aux besoins*

Il s'agit de la faculté des données à répondre aux *besoins réels*. Il n'est pas rare de voir des écrans surchargés d'informations qui n'apportent rien à personne. Il ne s'agit pas d'être restrictif lors de la conception du modèle de données, cependant il faut veiller à ne pas être trop « large » lors de la définition des informations à gérer. Inversement, il est extrêmement important de s'assurer que les attributs-clés et utiles pour l'entreprise soient bien prévus. Aussi, le niveau de détail des données est un élément à prendre en considération. Une donnée trop globale n'apportera pas la précision souhaitée à l'utilisation ; pour autant il est inutile d'être trop fin si ce niveau de finesse n'est pas requis.

3.2.1.3. *Facilité d'interprétation*

Il s'agit de qualifier la facilité d'interprétation de l'intitulé des attributs, d'une part, du contenu des attributs, d'autre part. Au-delà de la justesse, les données doivent être compréhensibles et ne laisser aucune ambiguïté quant à leur signification et leur

interprétation. Pour cela, les données doivent être lisibles, précises et simples tout en respectant les standards lorsqu'ils existent. Il faut se méfier en particulier des abréviations susceptibles d'être interprétées de différentes façons. Par exemple, l'article « Boîte de 10 – Diamètre 25 mm » désigne t-il une boîte de 10 écrous ou bien de 10 vis, sachant que le fabricant en question vend les deux articles ?

Le respect des standards a son importance non seulement pour la compréhension (par exemple la devise euro codifiée par EUR est reconnue par tout le monde) mais aussi pour les échanges de données entre applications, car les outils sous-jacents s'appuient sur les normes en vigueur.

3.2.2. *Accessibilité*

Par « accessibilité », nous entendons la disponibilité de l'information et la facilité avec laquelle on peut y accéder.

3.2.2.1. *Disponibilité*

L'information doit être disponible au moment où l'on en a besoin. Même si les systèmes de gestion « temps réel » sont de plus en plus répandus, il reste un certain nombre de données qui ne sont connues qu'avec un temps de retard par rapport à la réalité. C'est le cas de données qui résultent de traitements complexes appliqués à des volumes importants d'informations. Par exemple, le chiffre d'affaires consolidé des supermarchés d'une région ne sera généralement connu que le lendemain matin et non le soir même, car il faut « mouliner » toutes les données de chacun des supermarchés.

3.2.2.2. *Facilité d'accès*

Des données disponibles ne signifient pas pour autant qu'elles soient facilement accessibles. Les données-clés et régulièrement consultées doivent pouvoir être accessibles en quelques clics de souris et non par un enchaînement fastidieux

d'écrans. On touche là à l'ergonomie des applications, élément essentiel pour l'efficacité opérationnelle.

3.2.3. *Flexibilité*

La flexibilité des données traduit leur faculté d'évolution, de cohérence avec d'autres sources et de compréhension au-delà des frontières.

3.2.3.1. *Evolutivité*

Une donnée est évolutive si sa définition et sa codification lui permettent de vivre durablement sans remise en cause. La codification des données doit prendre en compte les volumes de créations futures d'enregistrements. Ainsi, si l'on décide de numéroter les articles automatiquement et séquentiellement, il faut estimer le nombre d'articles total qui seront contenus dans la base dans les années et décennies à venir afin de définir la tranche de numéros que l'on peut affecter aux articles. Supposons par exemple que l'on démarre une base articles avec 10 000 enregistrements et qu'il s'en crée environ 1 000 par an, on pourra définir une tranche de n° allant de 00001 à 99999, ce qui laisse une réserve de numéros de l'ordre d'une centaine d'années. L'utilisation de codes mnémoniques est à proscrire dans le cas de données volumineuses, car on arrive vite à saturation (voir section 7.2).

3.2.3.2. *Cohérence avec d'autres sources*

Les applications informatiques sont rarement isolées et les communications entre elles sont fréquentes. Il est donc important d'identifier les données partagées et les définir de façon cohérente. Cela facilite grandement l'élaboration des interfaces et améliore la fiabilité des informations communes. Comme déjà indiqué, il est recommandé de s'appuyer sur les normes en vigueur de codification chaque fois que cela est possible.

3.2.3.3. *Possibilités de traduction*

L'utilisation d'applications informatiques est de plus en plus internationale. Les données doivent donc être compréhensibles par tous. De nombreuses applications permettent désormais de manipuler les données dans plusieurs langues, ce qui suppose toutefois de les initialiser dans les différentes langues souhaitées.

3.2.4. *Sécurité*

La sécurité des informations est un aspect important à prendre en compte. Il s'agit de protéger l'information aussi bien contre les menaces accidentelles, que les attaques malveillantes. La sécurité recouvre la confidentialité, la fiabilité, la traçabilité et l'intégrité des données (voir paragraphe 1.5.3.4).

3.3. Conclusion

Principales caractéristiques de qualité des données	
Qualité du contenu	▪ Justesse ▪ Adéquation / besoins ▪ Compréhension
Accessibilité	▪ Disponibilité ▪ Facilité d'accès
Flexibilité	▪ Evolutivité ▪ Cohérence avec d'autres sources ▪ Traduction
Sécurité	▪ Respect de la confidentialité ▪ Fiabilité ▪ Traçabilité ▪ Intégrité

Tableau 3.1. *Principales caractéristiques de qualité des données*

Comme pour tout produit ou service, la qualité s'applique également aux données. Les seuls juges de la qualité des données sont les utilisateurs.

La justesse de l'information n'est pas le seul critère à prendre en compte, car d'autres dimensions aussi importantes que la présentation, l'accessibilité, la flexibilité jouent aussi un rôle essentiel dans la qualité des données.

Le tableau 3.1 présente les principales caractéristiques de qualité des données.

CHAPITRE 4

Causes et conséquences de la non-qualité des données

4.1. Les conséquences de la non-qualité des données

Dans beaucoup d'organisations, la plupart des personnes font leur travail du mieux qu'elles peuvent. Les efforts se concentrent sur la vente, la satisfaction des clients, le développement et l'amélioration des produits et services, la planification, la stratégie, le management ou encore la gestion des collaborateurs. En un mot, et c'est bien normal, la plupart des personnes cherchent à réaliser au mieux le travail qui leur est confié pour participer au succès de leur entreprise.

Si toutes ces activités nécessitent de l'information, donc des données, force est de constater que l'importance de celles-ci est souvent sous-estimée. Pour beaucoup, les données ne sont qu'un simple support et, outre quelques désagréments provoqués ici ou là par des « bugs », la qualité des données n'est pas considérée comme quelque chose d'essentiel. Le caractère immatériel des données est l'une des raisons de cette situation. Nous reviendrons de façon plus détaillée sur les raisons de la non-qualité à la section 4.3.

Pourtant, les conséquences de la non-qualité coûtent très cher. Et peuvent parfois conduire les entreprises à une situation critique. Les impacts négatifs de la non-qualité des données les plus couramment rencontrés sont les suivants :

– insatisfaction du client : les clients victimes de données incorrectes pardonnent difficilement ; une livraison qui arrive trop tard suite à une erreur sur la date de livraison ou une livraison partielle sont hélas des erreurs courantes ; avec internet, la sentence est encore plus rapide ; un internaute qui s'aperçoit que les données d'un site sont peu fiables ou incorrectes, comme par exemple des prix non actualisés, ne reviendra jamais sur le site ;

– non-conformité des chiffres publiés : des erreurs dans les données comptables peuvent entraîner de lourdes amendes et peines ; suite aux récentes affaires de falsification des comptes, toute erreur dans la publication des chiffres paraît désormais suspecte ;

– dévalorisation de l'image de l'entreprise : en conséquence du point précédent, des erreurs portant sur des données diffusées à l'extérieur de l'entreprise détériorent rapidement l'image de celle-ci et handicape son expansion commerciale ; de plus, l'impact boursier des sociétés cotées est immédiat, entraînant la chute rapide du cours de l'action ;

– perturbation du fonctionnement opérationnel : les donnés incorrectes au sein de l'entreprise ralentissent le fonctionnement opérationnel et rendent les décisions difficiles ou incertaines ; cela génère aussi du stress et de la démotivation chez les collaborateurs ; le surcroît de travail engendré par des erreurs de données n'est jamais le bienvenu ; la perte de confiance dans les informations manipulées n'est propice ni à l'initiative, ni à un climat sain ;

– erreurs de stratégie : définir une stratégie prend beaucoup plus de temps si les informations utilisées sont de mauvaise qualité ou tout simplement erronées ; comment réfléchir à l'avenir si la situation présente est mal maîtrisée et surtout mal

connue ? les dirigeants peuvent évidemment se fier à leur intuition, mais cela n'est généralement pas suffisant ;

– augmentation des coûts : les impacts énoncés précédemment engendrent évidemment des coûts ; les situations variant d'une entreprise à une autre, il est très difficile d'estimer le coût de la non-qualité ; ce coût caché peut cependant représenter jusqu'à 10 % du chiffres d'affaires, ce qui ampute de façon conséquente la rentabilité.

4.2. Quelques exemples de défauts de qualité des données

Ci-dessous sont présentés quelques exemples typiques de problèmes liés à la non-qualité des données.

4.2.1. *Doublons*

Les doublons correspondent à deux (ou davantage) enregistrements créés dans la base de données et décrivant une même réalité. Par exemple plusieurs « fiches client » sont stockés dans le SGBD pour un même client. Le problème des doublons vaut aussi pour d'autres types de données, par exemple fournisseurs ou articles. Ces cas relativement fréquents engendrent des perturbations opérationnelles. En effet, si plusieurs enregistrements correspondent au même client, lequel doit-on choisir lorsque l'on enregistre une commande ?

Le problème des doublons présente à terme des problèmes plus graves, car en cas de mise à jour, l'un des enregistrements est affecté mais pas l'autre, si bien que l'un possède des données correctes et l'autre non. Dans le cas d'une modification des coordonnées du client, un seul enregistrement sera désormais pourvu d'une adresse correcte. Et si l'on utilise un enregistrement incorrect, la marchandise a toutes les chances d'être livrée à la mauvaise adresse.

En outre, les doublons rendent difficiles les analyses par client. Le risque majeur est de se baser seulement sur une partie de l’activité du client et de ne pas le considérer, à tort, comme un client stratégique.

4.2.2. *Coordonnées des partenaires*

La gestion des coordonnées des partenaires clients ou fournisseurs (adresse, téléphone, fax, e-mail, contacts, etc.) est cruciale.

Il est cependant relativement fréquent de constater que ces données ne sont pas mises à jour avec toute la rigueur nécessaire. En conséquence, l’expédition de commandes à la mauvaise adresse n’est pas si rare.

Aussi, ces défauts de qualité engendrent des difficultés et de la perte de temps pour communiquer avec les partenaires. D’une certaine manière, cela se traduit par une connaissance affaiblie des partenaires, ce qui est négatif pour la réactivité et le business en général.

Par ailleurs, il faut veiller aux traitements automatiques : un fax ou un e-mail erroné peut retarder une commande d’achats urgente à un fournisseur et décaler la production s’il s’agit par exemple d’une matière première.

4.2.3. *Montants facturés*

Si ce cas fréquent peut avoir de multiples origines, les conséquences sont bien connues. En cas de sous-facturation, il y a perte de chiffres d’affaires si le client ne se manifeste pas. En cas de surfacturation, le mécontentement du client est immédiat.

4.2.4. *Disponibilité des produits*

Dans les entreprises, industrielles notamment, les flux physiques sont généralement tendus et les dates de livraison de marchandises doivent être fiables.

Or les dates de livraison sont de plus en plus déterminées par le système d'information qui prend en compte différentes contraintes, selon des règles parfois complexes. La justesse de cette donnée est donc liée à la qualité de multiples données et du modèle décrit dans le système d'information.

Lorsqu'une entreprise annonce une date de livraison à son client, ce dernier intègre cette date dans la planification de son activité (fabrication, vente).

Les dates annoncées ont donc des impacts directs et forts sur le business des clients. Or, pour diverses raisons, la justesse de cette information, primordiale pour les clients, n'est pas toujours au rendez-vous.

Les conséquences directes de cette non-fiabilité du délai de livraison sont donc :

– l'insatisfaction du client : doit-on continuer à travailler avec le fournisseur x si celui-ci annonce une date qui n'est pratiquement jamais respectée ?

– la perturbation de l'organisation de l'entreprise défaillante qui essaye de compenser l'insatisfaction du client en cherchant à le livrer au plus vite au détriment d'autres tâches ;

– la réputation et l'image de l'entreprise : le fournisseur incapable de satisfaire les délais annoncés est vite décrédibilisé ; ce phénomène est particulièrement sensible sur internet (on ne passe plus commande chez un fournisseur si le délai affiché n'est jamais respecté).

4.2.5. *Stocks*

Une gestion des stocks optimale nécessite que le système d'information soit en phase avec la réalité physique. Chaque article devrait avoir une quantité en stock dans le système d'information qui reflète la quantité réelle. Or là encore, il n'est pas rare de rencontrer des écarts entre les deux mondes.

Le risque est d'être soit en rupture de stock, soit de commander ou fabriquer à tort la marchandise. L'impact sur les coûts et donc la rentabilité de l'entreprise est évident.

4.2.6. *Indicateurs d'activité*

Les indicateurs d'activité tels que le chiffre d'affaires, le coût de revient des produits, la marge, le carnet de commandes, etc. s'appuient sur les données opérationnelles saisies dans l'activité quotidienne. Or si les informations élémentaires ne sont pas de bonne qualité, les indicateurs résultants ont de fortes chances de ne pas l'être non plus.

En conséquence, les analyses du management sont incorrectes et les décisions résultantes sont inappropriées. Le gap entre les décisions prises et celles qu'il aurait fallu prendre engendre rapidement une perte de crédibilité du management. Ce point est très sensible, car il impacte la qualité du pilotage et la pérennité de l'entreprise.

4.2.7. *Ergonomie*

Les problèmes d'ergonomie sont légion et peuvent être la source d'erreurs de saisie ou d'interprétation. Par exemple, certaines applications présentent des montants justes mais peu lisibles (pas de séparateurs de milliers par exemple). Cela augmente les risques d'erreurs dans la saisie ou l'interprétation de montants.

Aussi des écrans surchargés d'informations provoquent des difficultés d'adaptation et d'utilisation, avec là encore des risques d'erreurs de saisie. Ce type de défauts « cosmétiques » est généralement l'une des sources du mécontentement des utilisateurs et de rejet des systèmes informatiques.

Enfin, le contenu des informations en français uniquement peut présenter des difficultés à des utilisateurs étrangers. Même si les écrans sont traduits, il faut veiller que le contenu soit compréhensible par tous, sous peine d'engendrer des erreurs de saisie.

Cette liste pourrait facilement s'allonger, tant le champ des défauts de qualité de données est vaste. Remarquons que ces quelques cas mettent en évidence des anomalies de justesse de l'information, mais aussi d'autres natures de problèmes, comme par exemple des défauts d'ergonomie ou de traductions.

Comme déjà indiqué, la qualité ne repose pas uniquement sur la justesse de l'information, mais intègre d'autres facteurs tels que l'utilité ou la facilité d'interprétation.

4.3. Sources de non-qualité des données

Le nombre de données manipulées et stockées augmente sans cesse et les défauts de qualité deviennent difficiles à gérer si l'on ne combat pas ce problème à la source.

Les causes d'anomalies liées aux données sont nombreuses et variées. Globalement, on peut classer l'origine des anomalies dans l'une des catégories suivantes :

– sous-estimation de l'enjeu des données,
– conception,
– défaillance logicielle,
– initialisation des données.

4.3.1. *Sous-estimation de l'enjeu des données*

Au-delà de l'insuffisance de formation à la pratique de telle ou telle application, il y a très souvent une sous-estimation générale de l'importance des données. Plusieurs raisons expliquent ce constat :

– l'objectif d'une entreprise n'est pas de produire des données : bien que l'exercice des différents métiers nécessite des informations, le but premier est d'obtenir, selon le cas, un client satisfait, un niveau de vente supérieur, un meilleur produit, une meilleure marge, etc. ; les données ne sont donc pas au premier plan de l'activité et ne sont généralement pas perçues comme un élément essentiel de compétitivité ; grand nombre de managers considèrent que l'investissement dans une politique qualité des données est une pure perte de temps et d'argent ;

– le thème même de « la qualité des données » n'est guère attractif ; les ingénieurs préfèrent travailler sur des technologies de pointe, mettre au point des processus sophistiqués, plutôt que de s'intéresser au contenu, c'est-à-dire à l'information ; la qualité des données est très souvent le parent pauvre des projets informatiques ; pourtant, lorsqu'un problème lié aux données survient après le démarrage d'une application, personne n'est généralement enthousiaste pour procéder à des corrections et tout le monde admet qu'il aurait été préférable d'accorder en amont davantage d'importance aux données ;

– les données sont immatérielles et invisibles : ce côté abstrait met aussi l'information en arrière-plan des tâches concrètes telles que la négociation client, la production, les achats, le recrutement, etc. ;

– enfin de nombreux utilisateurs créent et gèrent des données utiles à leur job, mais peu se préoccupent de l'utilisation qui en sera faite en aval ; ce cloisonnement nuit considérablement à la qualité et à la fluidité de l'information.

4.3.2. *Défauts de conception*

Le modèle de données et les règles de gestion établies lors de l'étape de conception d'une application jouent un rôle déterminant dans la qualité des données. C'est à cette étape que l'on définit les données qui seront utilisées, par exemple, les informations que l'on souhaite utiliser dans une commande client, ainsi que les règles de gestion associées.

Aussi, le modèle de données et le travail effectué sur la définition et le contenu des informations n'est pas la seule source de non-qualité. La définition des flux et des processus joue également un rôle majeur. En effet, tout processus utilise des données en entrée et produit des données en sortie et il convient donc d'étudier à la fois le processus et les données.

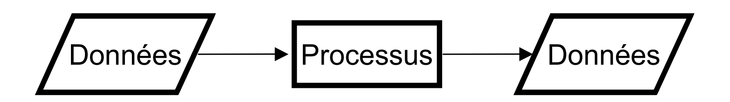


Figure 4.1. *Données et processus*

Les défauts de conception ne sont pas des exceptions et leur coût est important car la résolution *a posteriori* de ce type de problèmes impacte la chaîne de développement, tests, recette, etc.

C'est la raison pour laquelle il faut veiller à ce que les spécifications d'une application cible soient établies avec soin et répondent précisément aux besoins des utilisateurs. La compréhension des besoins est donc essentielle à cette étape.

4.3.3. *Anomalies du logiciel*

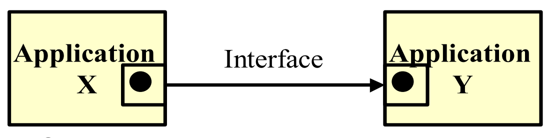
Les défauts du logiciel sont relativement courants au démarrage d'une application. Ils correspondent à un fonctionnement non conforme du logiciel par rapport aux

spécifications établies lors de l'étape de conception. Ce décalage provoque des anomalies qui se traduisent par des données incorrectes.

Cependant, si une majorité d'anomalies sont décelées les premiers mois suivant la mise en route d'une nouvelle application, des bugs peuvent survenir également par la suite. La raison en est simple : les applications et leur environnement évoluent avec le temps. Le paysage informationnel n'est généralement pas figé et évolue en permanence.

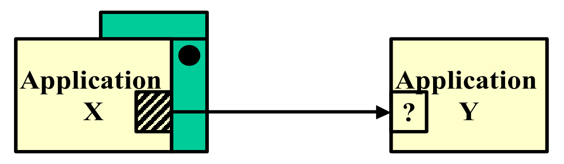
L'évolution d'une application donnée n'engendre jamais trop de problèmes en soi. Comme le montre la figure 4.2, ce sont souvent les impacts de l'évolution d'une application X sur une application Y qui sont sources d'anomalies.

Avant évolution de l'application X :



● Des données sont lues par l'interface dans une table de la base de données de X et transmises à l'application Y. L'interface sait où chercher ces données dans X et où les écrire dans Y.

Après évolution de l'application X :



L'application X évolue : certaines données sont désormais gérées dans une nouvelle table.

L'interface ne trouve plus les données dans la structure habituelle. L'application Y n'est plus alimentée.

Figure 4.2. *Impact de l'évolution d'une application sur une interface*

Pour illustrer ce schéma, prenons l'exemple d'un système de gestion maintenance industrielle permettant d'enregistrer et de

gérer les opérations de maintenance sur des équipements. Ce système communique initialement avec un système de suivi des coûts.

Chaque fois qu'un ordre de travail est réalisé sur un équipement, les données correspondantes sont saisies dans le système de gestion de maintenance. Puis, les informations permettant de calculer le coût du travail (pièces changées, temps de main d'œuvre) sont communiquées au système de suivi des coûts *via* une interface quotidienne.

Supposons qu'une refonte de ce système de gestion de maintenance déporte désormais les données de saisie des temps de main d'œuvre dans de nouvelles tables au sein d'un nouveau module dédié à la gestion des temps de travail. L'interface alimentant le système de gestion des coûts ne pourra donc plus véhiculer les éléments permettant le calcul des coûts de main d'œuvre, puisque les données temps de main d'œuvre sont désormais gérées dans d'autres tables. Si rien n'est fait, les coûts de maintenance seront donc inexacts et chuteront rapidement et anormalement.

Il est donc important de considérer les applications informatiques dans leur ensemble chaque fois qu'une évolution est effectuée.

4.3.4. *Initialisation des données*

L'initialisation des données dans un système de gestion revêt deux formes distinctes :

– les données saisies ou chargées au démarrage de l'application,

– les données saisies au fil de l'eau en régime permanent.

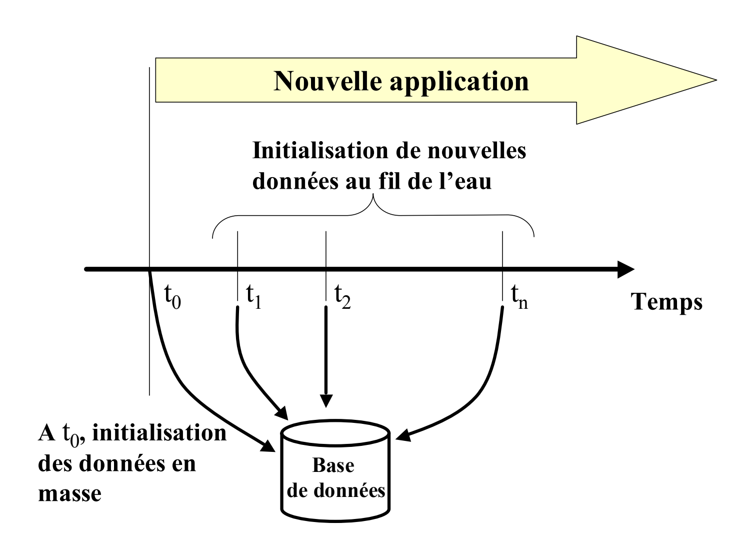


Figure 4.3. *Initialisation des données dans une application*

4.3.4.1. *L'initialisation des données au démarrage d'une application*

L'initialisation des données dans un nouveau système est une tâche cruciale. On peut véritablement l'assimiler à un transfert de patrimoine, car les données que l'on initialise dans le nouveau système, proviennent généralement d'un ancien système.

Or si les données de l'application d'origine ne sont pas de bonne qualité et que le transfert se fait sans nettoyage, l'application cible recueillera les mêmes défauts de qualité. Cela est particulièrement vrai pour les données stables, comme la base clients, fournisseurs ou encore articles. Ainsi, il n'est pas rare de transférer des doublons ou des données obsolètes d'une application à une autre.

Les tâches de contrôle des données avant initialisation en masse prennent du temps et cela mérite d'être pris en considération avec attention. Les investissements dans une nouvelle application ne méritent-ils pas d'avoir des données de qualité ?

Nous reviendrons en dernière partie de cet ouvrage sur la migration de données car c'est un point très sensible d'un point de vue qualité.

4.3.4.2. *La saisie des données au fil de l'eau*

Une fois l'application démarrée, de nombreux utilisateurs créent quotidiennement de nouvelles données. Les erreurs de saisie constituent la source principale des anomalies.

Une partie des erreurs de saisie est évitée car les applications prévoient des contrôles. Par exemple, dans la zone « unité d'achats » d'un article, il ne sera pas possible de saisir autre chose que l'une des unités stockées dans une table dédiée.

Unité	**Désignation**
kg	Kilogramme
L	Litre
m	Mètre
PCE	Pièce

Tableau 4.1. *Exemple de table des unités de mesure*

Cependant, rien n'empêchera de saisir « kg » à la place de « m ». Et c'est ici que le problème surgit. La machine ne peut pas tout contrôler. Plus exactement, elle ne peut contrôler que ce que l'on a prévu qu'elle contrôle.

Les anomalies couramment rencontrées résultent le plus souvent :

– d'erreurs de saisie involontaires : il peut s'agir d'une mauvaise manipulation ou d'un défaut d'attention de l'utilisateur qui ne saisit pas ce qu'il croit saisir ; même si les contrôles automatisés sont assez poussés pour arriver à un degré d'erreur voisin de zéro, il y a toujours une limite ; parmi les solutions permettant de limiter les erreurs de saisie, soulignons que les techniques de lecture code à barres sont d'une efficacité

impressionnante ; ce type de technologie est largement utilisé dans la grande distribution ;

– de la méconnaissance des règles de gestion de l'entreprise : typiquement, il s'agit de données incorrectes entrées dans le système d'information, non pas suite à une mauvaise manipulation, mais plutôt suite à une interprétation erronée de la signification des données saisies.

Prenons l'exemple d'un utilisateur venant de changer de service. Cet utilisateur est habilité à passer des commandes d'achats, comme il l'était déjà dans son ancien poste. Lors de la saisie d'une commande, un centre de coûts doit être saisi. Si cette personne ne s'en soucie pas, elle continue à saisir ses commandes d'achats exactement comme auparavant, et les nouvelles commandes d'achats seront imputées sur le centre de coût lié à son ancien service.

Dans ce cas précis, l'erreur est liée au fait que personne n'a informé cet utilisateur qu'il devait désormais imputer les achats sur un nouveau centre de coût. Il y a dans cet exemple à la fois un défaut de procédure et un problème de contrôle du système (car on pourrait imaginer un contrôle automatique du centre de coût, en fonction du rattachement au service).

La méconnaissance des règles de gestion traduit non seulement le *manque de formation* à l'utilisation de l'application informatique, mais surtout l'insuffisance d'information sur le fonctionnement même de l'entreprise.

La formation joue donc un rôle essentiel dans la prévention des anomalies. Outre la formation préalable au lancement d'une nouvelle application, des « piqûres de rappel » après démarrage sont indispensables.

4.4. Avantages compétitifs d'une politique qualité de l'information et des données

Nous l'avons vu dans les paragraphes précédents, l'absence de qualité des données handicape gravement la performance et l'image des entreprises. Inversement, nous présentons ici les avantages compétitifs que l'on peut tirer de la qualité de l'information et des données.

4.4.1. *Réactivité*

La qualité des données fluidifie les circuits d'information et entraîne globalement (et parfois de façon spectaculaire) une grande réactivité de l'entreprise. Quelle satisfaction pour un client de pouvoir se faire confirmer une date de livraison fiable et en temps réel ! Le monde moderne des affaires n'apprécie ni le flou, ni l'attente. Et c'est bien l'information juste, pertinente et communiquée au bon moment qui peut faire la différence entre plusieurs entreprises concurrentes. En résumé, la réactivité engendrée par la qualité des données est la clé de la *satisfaction du client* et de *l'image* de l'entreprise. Ce constat est d'ailleurs vérifié : les entreprises actuellement gagnantes sont celles qui misent sur leur système d'information et veillent à la qualité du contenu.

Prenons l'exemple d'une société dont le niveau des stocks enregistré dans le système d'information est rigoureusement identique au stock physique. Cette société saura se réapprovisionner au bon moment et ne risquera pas des ruptures de stocks, ni du sur-stockage résultant d'une mauvaise information. L'entreprise sera capable de livrer les clients comme prévu et sans surcoût parce qu'elle aura maîtrisé ses stocks.

Autre exemple : imaginons qu'un client passe une commande devant être livrée rapidement. Quel avantage de pouvoir le livrer à la date indiquée et à la bonne adresse sans cafouillage ! Cette

réactivité et la rigueur dans la gestion qui l'accompagne seront durablement appréciées par le client en question.

4.4.2. *Connaissance des clients*

La connaissance des clients et du marché peut être particulièrement améliorée par une politique qualité des données. Aujourd'hui, elle est devenue le nerf de la guerre : mieux connaître ses clients permet de tisser des liens forts et durables et garantit la pérennité de l'entreprise. Aucune organisation ne pourra survivre si elle ne connaît pas précisément ses clients et ses marchés.

L'une des clés de la connaissance des clients réside dans la gestion rigoureuse de la base de données clients. Avec une gestion efficace des coordonnées (adresses, téléphone, e-mail, fax…), on élimine le risque de communiquer « dans le vide » avec son client.

Avec des données justes comme le chiffre d'affaires et le détail des commandes réalisées avec un client donné (et non sur une multitude de doublons), les communications seront plus justes et en adéquation avec ses besoins. On diminue dans ce contexte le risque de considérer à tort le client comme étant de taille modeste, alors que la somme des doublons représente en fait un CA conséquent.

4.4.3. *Gestion opérationnelle optimisée*

La qualité de l'information améliore significativement les échanges entre services d'une même société. Le *travail est facilité* parce que les données manipulées sont plus fiables et complètement adaptées au business. Aussi, les données ne sont plus perçues comme un simple support de gestion, mais plutôt comme un précieux allié permettant d'optimiser à la fois son propre job et la performance globale de l'entreprise. Le temps

passé à corriger les données est largement diminué pour la plus grande *satisfaction des collaborateurs*.

Par exemple, le passage d'une commande d'achats à un fournisseur peut être très rapide si le prix de l'article pour ce fournisseur est stocké de façon fiable dans la base de données. Dans le cas contraire, on risque de ne pas optimiser l'achat. Pire, en cas de doute permanent sur la qualité des informations, les multiples coups de fil au service achats, voire aux différents fournisseurs potentiels, seront source de perte de temps et d'insatisfaction pour les approvisionneurs.

4.4.4. *Pertinence des décisions*

La confiance en l'information permet de décider plus vite et de façon plus juste. La conduite de l'entreprise est donc davantage sécurisée avec des données pertinentes et fiables. Les opportunités et les pièges sont mieux appréhendés. Même si la prise de décision n'est jamais simple, le flou qui peut apparaître avec des données peu fiables est effacé. Le cap est donc mieux maintenu et les tâches d'exécution qui en découlent se font avec confiance et moins d'à-coups. Cette visibilité accrue permet aussi de définir *une stratégie plus claire et moins hasardeuse*.

Prenons deux exemples concrets :

– cas du calcul du coût de revient d'un produit fabriqué : une entreprise dispose d'un logiciel permettant de calculer le coût de revient des produits finis ; de nombreux paramètres interviennent dans le calcul : sa nomenclature (c'est-à-dire la constitution du produit fabriqué à partir de ses composants), sa recette décrivant les étapes de la fabrication (analogue à une recette de cuisine), les temps et coûts de main d'œuvre, etc. ; la qualité de chacun des paramètres donnera lieu à un calcul du coût de revient fiable. Et chaque produit pourra réellement être considéré à sa juste valeur et apprécié comme étant un produit vedette ou au contraire un foyer de perte ; on ne risquera pas

l'arrêt à tort d'un produit rentable ou inversement la mise en avant exagérée d'un produit en fait trop coûteux ;

– cas des clients stratégiques : si la base de données est saine, les clients stratégiques seront très vite identifiés par leur chiffres d'affaires, leur marge, etc. ; on ne risquera pas de faire l'impasse sur certains clients qui seraient enregistrés plusieurs fois et qui n'apparaîtraient pas de façon consolidée comme des clients importants.

Certes, au-delà des analyses chiffrées délivrées par la machine, le savoir et l'expérience des responsables devrait mettre en défaut les analyses en question si elles ne sont pas justes. Et les erreurs devraient être démasquées rapidement.

Cela est vrai si les écarts sont importants. S'ils le sont moins, il n'est pas évident qu'une personne découvre des anomalies au vu des analyses. Or une somme de petites anomalies non décelables peut cacher un problème de taille.

Ces deux cas mettent donc clairement en évidence l'importance de la qualité de l'information dans les systèmes opérationnels pour analyser et prendre de bonnes décisions.

4.4.5. *Maîtrise des coûts*

L'impact sur les coûts des avantages cités précédemment est évident : à ressource égale, il y a moins de temps perdu, moins de travail de correction de données et plus d'efficacité dans les tâches opérationnelles. A cela s'ajoutent des recettes supplémentaires dues à une satisfaction du client accrue, de meilleures décisions, une stratégie plus pertinente.

Au final, la qualité des données se traduit par des économies de coûts. La réactivité, la maîtrise des clients, l'optimisation des tâches, la pertinence des décisions vont dans le sens d'une baisse des coûts.

4.5. Conclusion

Les données ne sont généralement pas considérées comme vitales au fonctionnement des organisations. Pourtant, les défauts de qualité des données pèsent fortement sur la profitabilité et la compétitivité des entreprises et impactent directement ou indirectement la satisfaction des clients.

Si les sources de non-qualité sont nombreuses, la méconnaissance des enjeux et des impacts de la qualité des données sur le business est de loin la raison majeure de cette négligence. D'où l'importance des actions de sensibilisation, de communication et de formation au rôle-clé des données dans l'entreprise.

Avec ce chapitre s'achève la deuxième partie de l'ouvrage. Il est maintenant clair que la qualité des données est indispensable à toute organisation souhaitant « surperformer » dans son secteur. Il est même sage de prévenir que le choix est clair : sacrifier la qualité de l'information plongera tôt ou tard l'entreprise dans une situation d'échec ; développer au contraire un plan qualité de l'information favorisera son succès.

TROISIÈME PARTIE

Comment améliorer la qualité des données ?

CHAPITRE 5

L'évaluation de la qualité des données

La qualité des données s'apprécie difficilement et l'avis des utilisateurs est généralement insuffisant et trop imprécis pour se faire une idée exacte de l'état des données.

C'est pourquoi il est utile de se doter de moyens de contrôle moins subjectifs. Aussi, il est important de communiquer et d'expliquer les objectifs des actions de contrôle, à savoir déceler les anomalies pour en corriger la source et améliorer la performance globale.

Il n'est pas nécessaire d'utiliser des techniques sophistiquées pour mesurer le niveau de qualité des données. Les méthodes doivent être adaptées au niveau de maturité des organisations. En termes de responsabilité et de moyens humains, l'évaluation de la qualité est menée conjointement par la direction des systèmes d'information et les principaux acteurs des domaines fonctionnels impactés.

5.1. Mesure du taux de données erronées

Il s'agit de l'évaluation la plus simple qui puisse être effectuée. L'idée est de mesurer le taux de données incorrectes

d'un échantillon représentatif. Il est recommandé de procéder à des mesures que s'il y a un réel enjeu en termes de business. Il est donc important de définir préalablement ce que l'on veut évaluer et dans quel but.

Plus qu'une série de mesures, il s'agit d'une démarche comportant les étapes décrites sur la figure 5.1.

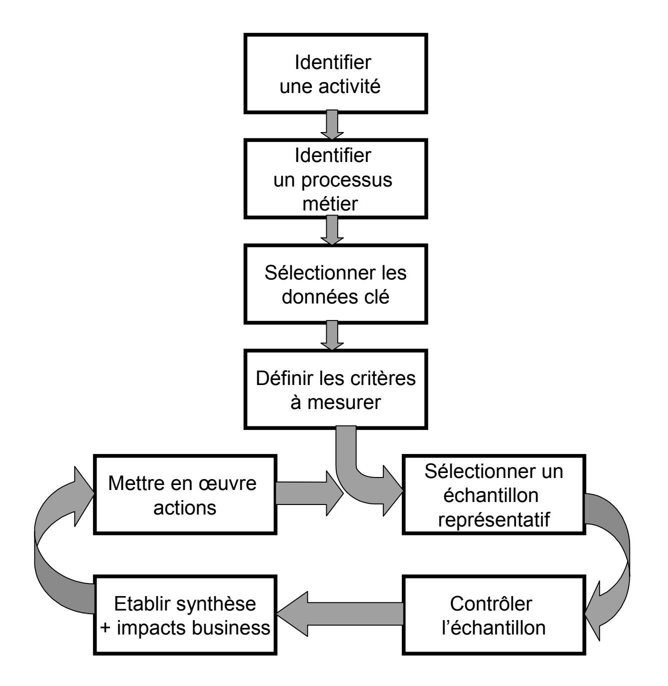


Figure 5.1. *Démarche d'évaluation de la qualité*

5.1.1. *Identifier une activité*

Il s'agit de choisir une activité-clé, au sein de laquelle des anomalies sont constatées. Il peut s'agir d'une activité commerciale, logistique, production, achats, comptable, etc. Généralement, le constat que la chaîne informatique en place ne répond pas aux besoins des utilisateurs est une raison majeure de non-qualité.

L'évaluation de l'enjeu de l'activité est primordiale : il ne faut pas perdre de temps à chercher des améliorations dans des activités qui ne sont pas stratégiques. Il est difficile de préciser ce qu'est une activité stratégique, car cela dépend évidemment de l'entreprise et du marché sur lequel elle opère. Néanmoins, les activités ayant un impact fort sur la satisfaction client doivent être privilégiées (vente et logistique en particulier).

Plusieurs activités peuvent être identifiées, mais encore une fois, il est conseillé de ne pas se disperser et de fixer des priorités. Il est même parfois raisonnable de ne pas chercher à être puriste dans la recherche de qualité. Si l'impact de la qualité sur le business est négligeable, il est inutile d'y consacrer du temps et de dépenser de l'argent.

Une fois l'activité à auditer identifiée, c'est avec l'accord du manager que l'analyse doit être menée. Dans certaines entreprises, il est intéressant de constater que ce sont les managers d'activité eux-mêmes qui sont demandeurs d'un programme d'amélioration de la qualité des données. L'initiative n'est donc pas toujours, et c'est une bonne chose, à l'initiative de la direction des systèmes d'information.

5.1.2. *Identifier un processus métier*

Au sein de l'activité précédemment définie, un processus critique en termes de données est identifié.

L'avis du manager de l'activité sélectionnée est essentiel. Cet avis, recoupé des informations provenant de l'équipe maintenance informatique (qui trace toutes les anomalies et autres bugs), permet de bien cerner le point faible de l'activité.

Par exemple, on pourra s'intéresser au processus « enregistrement d'une commande de vente » au sein de l'activité administration des ventes.

5.1.3. *Sélectionner les données*

Il s'agit d'identifier les objets de gestion et les attributs (ou zones) jouant un rôle-clé dans le processus.

Par exemple, on sélectionnera toutes les zones (saisies et calculées) de la commande de vente, s'il s'avère que la qualité de toutes ces informations est insuffisante.

La participation des acteurs principaux du processus analysé est indispensable. La connaissance métier est requise là encore pour connaître ce qui est vital en termes de données dans le déroulement de tel ou tel processus.

5.1.4. *Définir les critères à mesurer*

L'objectif est ici de définir les critères à mesurer vis-à-vis des zones sélectionnées.

Les critères à prendre en compte peuvent varier d'un contexte à un autre. Il est néanmoins souvent intéressant de mesurer le taux d'enregistrements corrects, le taux d'enregistrements incorrects mais utilisables et le taux d'enregistrements incorrects non utilisables.

Ces critères très simples d'évaluation de la qualité constituent une excellente approche. Il ne faut pas forcément essayer de mesurer des choses complexes. La simplicité et le pragmatisme sont en effet les meilleurs alliés de la qualité.

5.1.5. *Sélectionner un échantillon représentatif*

L'idée est d'apprécier la qualité non pas sur la totalité de la base de données, mais sur un échantillon représentatif. La raison majeure est que l'analyse de la qualité est essentiellement une analyse humaine. Par conséquent, il est impensable et

impossible de chercher les défauts de qualité sur des dizaines de milliers d'enregistrements.

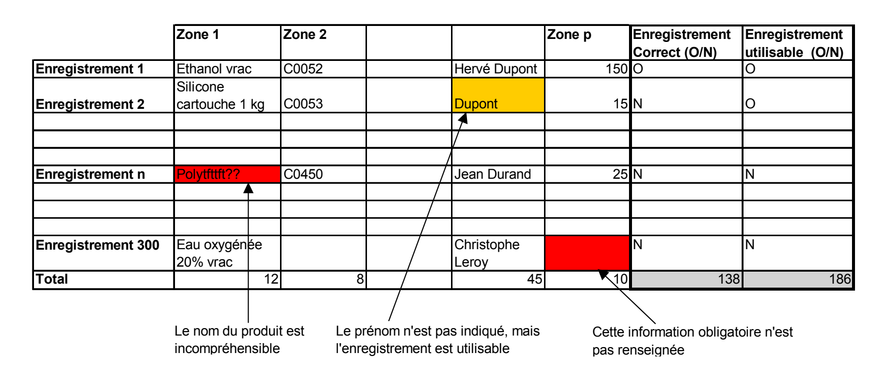
On sélectionnera par exemple, toutes les commandes de vente des trente derniers jours. L'échantillon doit être choisi pour sa pertinence et sa représentativité.

5.1.6. *Contrôler l'échantillon*

Cette étape consiste à contrôler l'échantillon et à reporter les erreurs pour chacun des critères retenus.

Le contrôle de l'échantillon consiste à vérifier les zones sélectionnées de chaque enregistrement de l'échantillon et à indiquer si elles répondent aux critères définis. Ce travail peut paraître contraignant et fastidieux, mais si les critères de mesure sont très précis, cela peut aller relativement vite (de l'ordre d'un homme-jour pour 50 enregistrements).

D'un point de vue pratique, les enregistrements sélectionnés doivent être présentés sous la forme d'un tableau Excel, comme le montre l'exemple de la figure 5.2.



	Zone 1	Zone 2			Zone p	Enregistrement Correct (O/N)	Enregistrement utilisable (O/N)
Enregistrement 1	Ethanol vrac	C0052		Hervé Dupont	150	O	O
Enregistrement 2	Silicone cartouche 1 kg	C0053		Dupont	15	N	O
Enregistrement n	Polytfttft??	C0450		Jean Durand	25	N	N
Enregistrement 300	Eau oxygénée 20% vrac			Christophe Leroy		N	N
Total	12	8		45	10	138	186

Figure 5.2. *Exemple de contrôle des enregistrements*

5.1.7. *Etablir une synthèse des résultats*

Une synthèse très claire des résultats permet à la fois de faire ressortir les données présentant des anomalies majeures et de communiquer au sein de l'entreprise.

Cette synthèse doit également mettre en évidence les principaux impacts sur le business. Une représentation graphique (diagramme en bâtons ou autre) est certainement l'outil le plus approprié pour communiquer les résultats.

Dans l'exemple précédent, sur un échantillon de 300 enregistrements, il y a 138 enregistrements corrects, 186 enregistrements utilisables, 114 enregistrements défectueux inutilisables.

Nb enregistrements	**300**
% corrects	**46 %**
% utilisables	**62 %**
% inutilisables	**38 %**

Tableau 5.1. *Exemple de synthèse des résultats de contrôle*

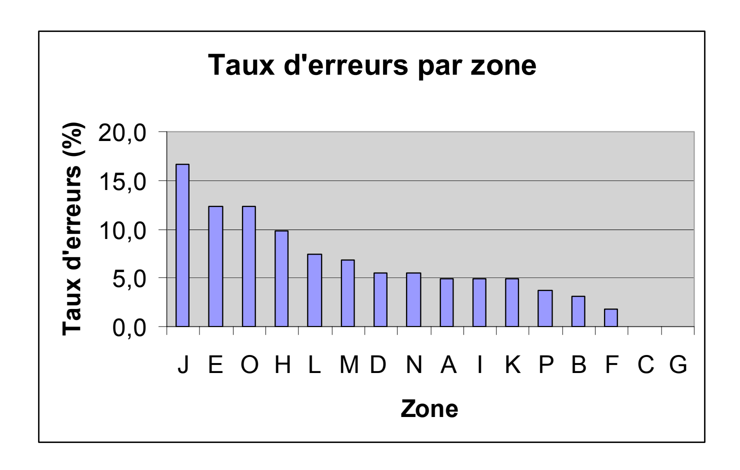


Figure 5.3. *Exemple de graphique « taux d'erreurs par zone »*

Il est intéressant de publier également les taux d’erreurs par zone (ou attribut) pour mettre en évidence les zones présentant le plus d’anomalies.

Ce type de tableau doit être accompagné d’une note écrite mettant l’accent sur les impacts « business ».

5.1.8. *Définir un plan d’actions*

Suite à l’analyse précédente, un plan d’actions permettant de résoudre les anomalies doit être mis en œuvre. La résolution des anomalies est décrite au chapitre 6.

5.1.9. *Mesurer l’impact des actions*

Des mesures après correction doivent être menées (voir chapitre 6).

5.2. Détection d’erreurs au cours d’un processus

La mesure du taux d’erreurs décrite au paragraphe précédent est basée sur une photographie de la base de données à un instant *t* et permet d’évaluer rapidement et simplement le niveau de qualité des données.

Il peut être intéressant d’aller plus loin et de déterminer à quelle étape d’un processus les erreurs se produisent. Par exemple, on étudiera l’évolution des données au cours du processus de saisie de commandes clients. Une assistante commerciale crée une commande client (étape 1) ; une modification de la quantité est ensuite saisie suite à appel du client (étape 2) ; un programme automatisé est alors exécuté pour déterminer la date de livraison (étape 3), etc.

Processus comprenant n étapes

Etape 1	→	Etape 2	→	Etape 3	→	Etape n
Màj 1		Màj 2		Màj 3		Màj n

Base de données

Figure 5.4. *Mises à jour successives de la base de données*

A partir des enregistrements d'un échantillon représentatif, un travail d'analyse est effectué. Ce travail peut être effectué sur un tableau Excel des enregistrements de l'échantillon sélectionné.

Il y a donc besoin d'un travail préparatoire pour extraire les données à chaque étape du processus. Ce travail peut se faire en fin de processus si l'application permet de mémoriser les différentes mises à jour successives.

	Etape 1	Etape 2	Etape 3	…	Etape n
N° Commande	160200301	160200301	160200301		160200301
Statut	C	V	V		L
Date livraison	11/12/04	11/12/04	11/12/04		11/12/04
Article	85410	85410	85410		85410
Quantité	1.250	1,250	1.250		1.250
Unité	kg	TO	TO		TO
Client	145223	14522	14522		14522

La quantité est exprimée en tonnes et non en kg, mais la conversion est correcte, car la quantité est divisée par 1000

Erreur : le code client est tronqué. La commande est donc prise pour un mauvais client

Erreur : la virgule (séparateur de décimales) est remplacée par un point (séparateur de milliers). La quantité est donc multipliée par 1000 !

Figure 5.5. *Analyse des modifications d'enregistrements*

L’analyse des modifications de la base de données permet alors d’identifier les anomalies à chacune des étapes du processus étudié et d’agir avec précision là où les erreurs se produisent.

La mise en œuvre de ce type d’analyse est relativement long, car il nécessite d’interroger le journal de la base de données. Mais il est aussi plus riche d’enseignements car il permet de comprendre à quelle étape d’un processus il y a problème. Le temps passé à collecter l’information est largement compensé par le temps économisé lors de la résolution du problème.

Une synthèse par étape du processus apporte l’éclairage nécessaire à une bonne analyse, voir figure 5.6.

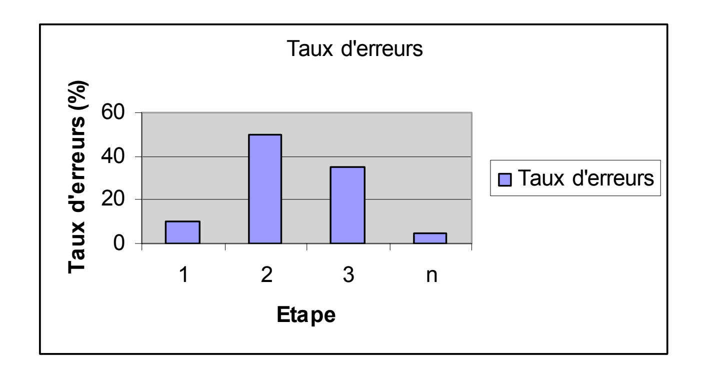


Figure 5.6. *Exemple de synthèse des résultats de contrôle par étape*

Dans cet exemple, la majorité des erreurs se produisent dans les étapes 2 et 3.

5.3. Contrôle statistique

Le contrôle statistique permet d’établir une base fiable de taux d’erreurs et de détecter si ce taux augmente anormalement. Le principe est d’établir des mesures périodiques du taux d’erreurs et de les comparer avec le niveau d’exigence que l’on se fixe.

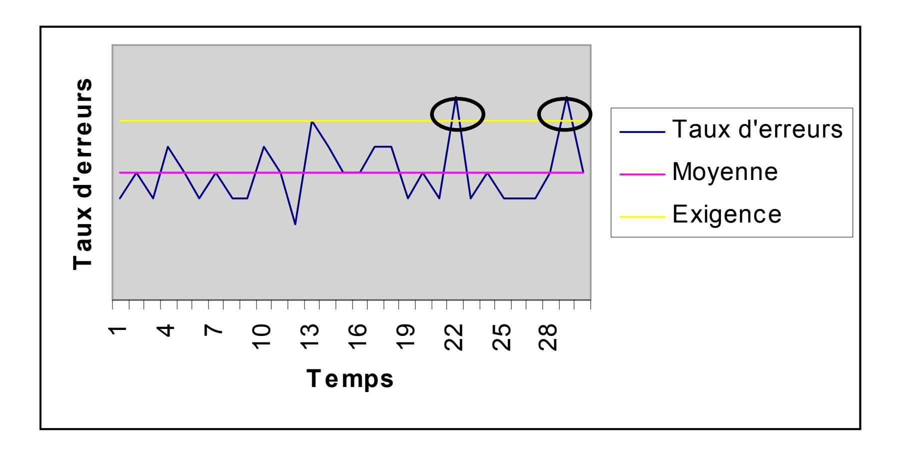


Figure 5.7. *Exemple de contrôle statistique*

Dans cet exemple, on constate que le niveau d'exigence est dépassé à deux reprises en fin de mois. On note donc une détérioration de la qualité des données et l'on doit en rechercher les causes pour les éliminer.

Le contrôle statistique permet d'anticiper les problèmes car il met en évidence assez tôt des niveaux d'anomalies inhabituels. C'est un outil d'alerte particulièrement efficace, qui peut être utilisé très simplement comme dans l'exemple précédent. Evidemment, l'une des difficultés de cette méthode est de définir le bon niveau d'exigence. La meilleure façon de procéder est de fixer l'exigence avec le responsable du domaine analysé. Par exemple, si l'on mesure le taux d'erreurs dans le processus de prise de commande, alors il convient de définir l'exigence avec le responsable du service clients.

Aussi, nous n'aborderons pas dans cet ouvrage les techniques de contrôle statistique plus sophistiquées, qui bien qu'efficaces, sont plus difficiles à mettre en œuvre. De nombreux ouvrages traitant des probabilités et statistiques fournissent des modèles de prévision et il peut être intéressant de les lire pour aller plus loin.

5.4. Conclusion

Quelques outils très simples à mettre en œuvre permettent d’évaluer avec précision le niveau de qualité des données. A partir de ces informations, il convient de procéder à des améliorations puis d’effectuer à nouveau des mesures et de recommencer le cycle jusqu’à disparition des anomalies. Le chapitre suivant décrit de façon détaillée le processus d’amélioration.

CHAPITRE 6

L'élimination des défaillances

6.1. L'approche globale

Dans la plupart des organisations, la majeure partie des tâches des équipes de support ou de maintenance informatique consiste à gérer et corriger les anomalies au fil de l'eau.

Cette approche fonctionne bien et soulage les utilisateurs qui demandent à être « dépannés » au plus vite. Elle est néanmoins parfois insuffisante pour garantir un niveau de qualité durable. Il est en effet conseillé d'aller plus loin dans l'anticipation des problèmes.

Grâce aux analyses présentées au chapitre 5, il est relativement facile de mesurer le niveau de qualité des données d'une activité ou d'un processus. A partir de cette analyse, l'idée est de se focaliser sur les problèmes critiques en termes de business. La méthode recommandée est alors de mener ponctuellement des projets d'amélioration de la qualité des données. Par exemple, un projet pourra être l'amélioration de la prise de commande clients ; un autre projet pourra être l'amélioration du processus de facturation, etc. Ces projets sont menés par la direction des systèmes d'information et impliquent

fortement les responsables et les équipes du ou des domaines fonctionnels concernés.

Cette approche diffère donc de la maintenance classique qui consiste à corriger les bugs au fur et à mesure qu'ils sont détectés. Aussi, la méthode proposée est nettement plus globale car elle prend en considération à la fois les aspects techniques et organisationnels. Ce qui déclenche un tel projet n'est pas la constatation d'un dysfonctionnement ponctuel, mais plutôt la mise en évidence d'un niveau insuffisant de qualité des données au sein d'une activité ou d'un processus.

La méthode préconisée peut être schématisée par la figure 6.1.

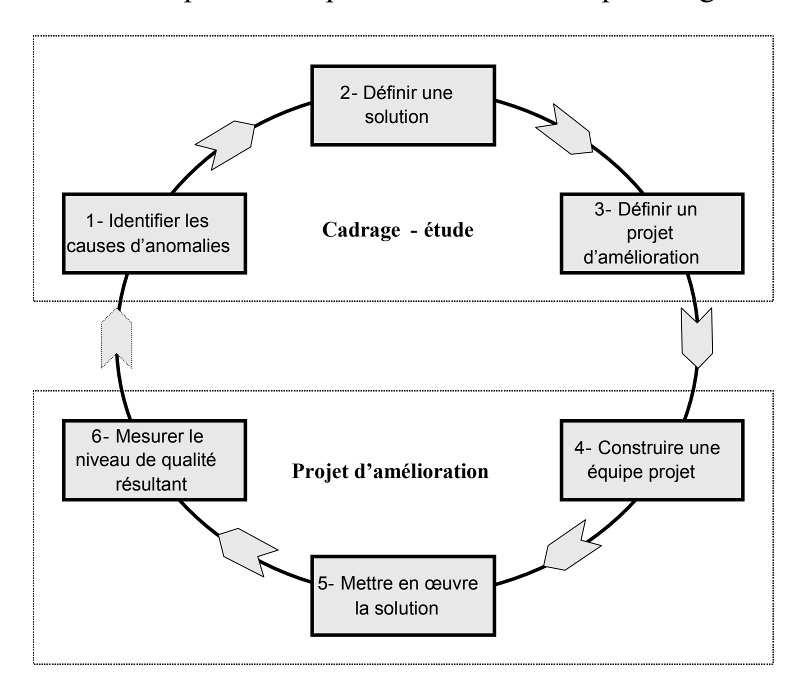


Figure 6.1. *Méthode d'élimination des défaillances*

Les paragraphes suivants détaillent chacune des étapes de la démarche.

6.2. Identifier les causes d'anomalies

Suite à la mise en évidence de défauts de qualité de données (voir chapitre 5), l'objectif de cette étape est de chercher la ou les sources du problème posé. L'analyse du processus doit être effectuée de façon méticuleuse, en s'aidant d'extractions de la base de données. Ce travail s'apparente à celui d'une enquête. Si la direction des systèmes d'information participe largement à l'identification des causes, il est indispensable d'impliquer les personnes travaillant directement au cœur du processus et de comprendre ce qu'elles font et comment. Les erreurs proviennent très souvent d'une mauvaise utilisation du système et pas uniquement d'anomalies informatiques.

Lorsque un certain nombre d'éléments sont collectés, l'équipe dédiée à cette étude se réunit pour discuter et partager les causes du problème. Plusieurs réunions de ce type sont nécessaires pour converger et lister les sources d'anomalies.

6.3. Définir une solution

Lorsque les causes sont bien identifiées, une ou plusieurs solutions doivent être définies. Il s'agit d'étudier et de décrire les actions à mettre en œuvre pour remédier aux anomalies décelées. Rappelons qu'il est plus rentable de privilégier les actions simples, plutôt que des modifications pointues, mais complexes à mettre en œuvre. Il faut aussi rester pragmatique et adopter une démarche préventive plutôt que corrective.

La solution peut comporter plusieurs aspects :

– technique : par exemple, modifier tel programme ou mettre en œuvre tel module d'un progiciel ; ou encore mettre en œuvre un système de saisie code à barres pour limiter les erreurs de saisie ;

– organisationnel : par exemple, redéfinir un circuit de validation ;

– formation : par exemple, prévoir un stage de formation à la prise de commandes.

Si plusieurs solutions sont proposées, il convient de trancher avant le lancement du projet en mettant en évidence avantages et inconvénients de chacune d'elles. L'estimation des coûts doit bien entendu faire partie des critères de décision.

Le choix de la solution à mettre en œuvre est bien sûr documenté. C'est en quelque sorte le cahier des charges du projet d'amélioration.

6.4. Lancer un projet d'amélioration

Lorsque la solution est définie, le projet d'amélioration peut être lancé. L'objectif du projet est issu directement de l'étape précédente. Par exemple : diviser par deux le taux d'erreurs dans la saisie des quantités lors de la prise de commande de vente, en mettant en œuvre la solution définie précédemment.

Un tel projet ne doit être lancé que s'il y a un réel besoin et un enjeu significatif pour l'entreprise. Il est donc recommandé de bien s'assurer que le processus concerné joue un rôle-clé dans la compétitivité de l'entreprise. Il faut aussi évaluer les chances d'aboutir à un résultat positif.

Le projet doit être porté par un propriétaire. Par exemple, si l'objectif est de diminuer le taux d'erreurs dans un processus achat, le propriétaire pourra être le chef de service achats. Le propriétaire est en quelque sorte un client ayant un besoin d'amélioration de la qualité de l'information au sein de son domaine.

Le budget du projet est estimé à cette étape. Cette tâche est facilitée, car l'identification des causes d'anomalie et le choix de la solution à mettre en œuvre ont été effectués en amont. Disons que le coût d'un tel projet dépend de la gravité du problème et

de l'ambition que l'on se fixe. L'amélioration de la qualité peut cependant être construite en plusieurs étapes ou sous-projets de façon à étaler les coûts dans le temps et décider après chaque étape si l'on poursuit le travail d'amélioration ou non.

6.5. Construire l'équipe projet

Tout d'abord, un leader doit être nommé. Les capacités de leadership, d'organisation, de communication sont essentielles pour ce rôle. Le leader est une personne appartenant à la direction des systèmes d'information (équipe maintenance applicative par exemple).

Le leader constitue alors une équipe, composée en fonction de la problématique et de la solution choisie. Une charte, claire et simple doit enfin être écrite avant le lancement du projet. Cette charte doit comprendre au minimum :

– la description du problème : faits (appuyés par des mesures), impacts sur le business ;

– l'objectif du projet. Par exemple, diviser le taux d'erreurs par trois en deux mois dans le processus de prise de commande, par la mise en place d'une solution X décrite au cahier des charges ;

– le fonctionnement de l'équipe : qui fait quoi, quand et comment ; noter que l'équipe ne doit pas nécessairement être à temps plein sur le projet ; il faut spécifier clairement la charge attendue pour chaque membre, par exemple *x* heures par jour ; le leader aura en charge le management de l'équipe, le suivi et la publication des résultats, l'animation de réunion d'avancement d'avancement et de « debriefing » ;

– le premier travail du leader du projet est d'établir un planning détaillé des tâches de l'équipe ; ce travail de planification est tout à fait analogue à ce qui se fait sur les projets de construction de systèmes d'information.

6.6. Détailler et mettre en œuvre la solution

La mise en œuvre de la solution est menée par l'équipe projet. Une analyse détaillée de la solution définie à la section 6.3 est d'abord effectuée.

Si la solution comporte un volet technique, il y a besoin de procéder à des tests et à des formations avant la mise en œuvre de l'amélioration technique (nouveaux écrans, nouveau programme…). Une fois testée, la solution technique peut être implémentée. Si au contraire la solution repose davantage sur des modifications organisationnelles, des actions de conduite du changement sont menées.

Quelle que soit la nature de la solution, il est primordial de s'assurer que celle-ci est bien comprise et maîtrisée des personnes impactées. La communication est la clé de la réussite de ce type de projet.

Il faut aussi prévoir d'accompagner les personnes dans leur travail après implémentation de la solution. Il ne faut surtout pas « laisser faire » à ce stade, mais au contraire être présent sur le terrain.

6.7. Mesurer le niveau de qualité résultant

La meilleure façon de contrôler que la solution retenue fonctionne est de procéder à des mesures du niveau de qualité des données après mise en œuvre de la solution. Ces mesures sont effectuées sur plusieurs jours, voire plusieurs semaines, pour pouvoir en tirer des conclusions fiables. La méthode permettant de mesurer le niveau de qualité des données est décrite au chapitre 5.

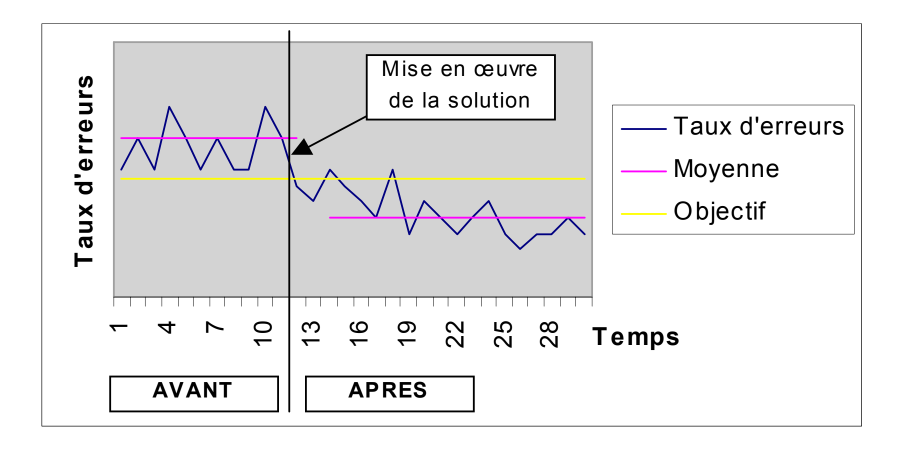


Figure 6.2. *Mesure du taux d'erreurs avant et après mise en œuvre de la solution*

Dans l'exemple du graphique précédent, la mise en œuvre de la solution engendre une chute du taux d'erreurs. La moyenne de ce taux sur une période significative après mise en œuvre de la solution est nettement en dessous de l'objectif. Le projet est donc un succès.

Une fois l'objectif fixé atteint, le projet peut être clos. La communication des résultats est effectuée avec des données chiffrées et des graphiques simples.

Si l'objectif n'est pas atteint, alors il faut recommencer le cycle, en modifiant le cas échéant le contenu des différentes étapes (revoir l'équipe, affiner la recherche des causes, identifier d'autres solutions, etc.).

6.8. Conclusion

Lorsqu'ils sont bien menés, les projets d'amélioration de la qualité des données sont très efficaces. Des objectifs raisonnables, une équipe motivée (non nécessairement à temps plein), des mesures fiables permettent d'obtenir assez rapidement des résultats probants.

CHAPITRE 7

Les bonnes pratiques de Data Management

Ce chapitre a pour objectif de définir les bonnes pratiques à adopter lors de la définition, la création et la modification des données. Ces recommandations contribuent à l'obtention d'informations de qualité.

Elles s'appliquent pour la plupart aussi bien à des applications en phase de conception, qu'à des applications démarrées depuis longtemps. Il n'est jamais trop tard pour mettre en œuvre de bonnes pratiques !

7.1. La compréhension des besoins

La bonne compréhension des besoins des utilisateurs est le point de départ de toute activité de conception ou de maintenance des systèmes d'information. Il est essentiel de comprendre et d'analyser de façon détaillée et avec pragmatisme le fondement de tel ou tel processus.

L'analyse et la formalisation des besoins est une étape-clé qu'il ne faut pas chercher à minimiser, car un besoin mal compris est catastrophique en termes de coût. Il ne faut surtout pas être avare de schémas, diagrammes de flux ou modèles

graphiques pour formaliser et valider les besoins. La plupart des méthodes de conception proposent toute une panoplie de modèles qui facilitent cette étape. Ainsi le modèle entité-relation (voir paragraphe 1.5.3.1) est indispensable à la formalisation des objets de gestion à gérer.

L'une des difficultés réside dans le dialogue entre deux populations distinctes :

– les utilisateurs : le chef comptable, le responsable commercial, le contrôleur de gestion, le directeur d'usine…

– les informaticiens : le développeur, l'analyste, l'ingénieur chargé d'études, le chef de projet, le consultant…

De façon schématique et simplifiée, les utilisateurs ont un savoir-faire, une connaissance poussée de leur domaine et ont des besoins qui ne sont pas toujours formellement exprimés. De leur côté, les informaticiens maîtrisent des outils et des techniques et raisonnent à partir d'informations clairement définies. Ce ne sont généralement pas des spécialistes métier.

C'est pourquoi les informaticiens ne doivent pas être avares de questions et doivent reformuler ce qu'ils entendent et comprennent. Outre la compétence informatique, l'une des qualités les plus recherchées est certainement l'aptitude à communiquer (comprendre l'interlocuteur), idéalement associée à une ou plusieurs compétences métier (la finance, le contrôle de gestion, la production…). Il devient de plus en plus nécessaire de parler le même langage que les utilisateurs.

Voici quelques principes importants à retenir, lors de la conception ou la maintenance d'une application :

– toujours avoir la vision globale d'un processus et se focaliser sur ce qui est important d'un point de vue business ;

– chercher à simplifier et éviter la construction d'usines à gaz ; si la solution devient vraiment trop complexe, c'est que l'on fait fausse route ;

– éliminer les étapes sans valeur ajoutée ;

– identifier les cas d'erreurs possibles et les mettre en évidence de façon compréhensible ;

– penser « données » dès la conception de la solution ;

– consacrer du temps à la documentation.

7.2. La codification des données

Les objets de gestion définis dans les modèles de données sont porteurs d'attributs (que l'on retrouve sous forme de zones sur les écrans). La codification consiste à :

– définir le format : nombre de caractères et type (numérique/alphanumérique),

– définir les valeurs possibles.

La codification doit être effectuée avec soin, car elle conditionne la bonne utilisation de l'application. Il est recommandé de définir les codes en considérant l'évolution future de l'application.

Voici quelques recommandations :

– s'appuyer sur les normes : il faut systématiquement rechercher si une norme ISO existe pour la codification d'une donnée ; par exemple, pour les devises, on utilisera la codification ISO 4217 des devises (EUR, USD…) ; pour les unités de mesures, on se basera sur les normes ISO 31 et ISO 1000 ; l'utilisation des normes garantit un langage commun à la plupart des applications informatiques et facilite les échanges entre systèmes ; bien sûr, tout n'est pas normalisé et il faut la plupart du temps inventer des codes, mais la recherche d'une norme doit être un réflexe ;

– porter une grande attention aux adresses : veiller à bien structurer les adresses (n° dans la rue, nom rue, code postal, boîte postale, ville…) ; faire attention en particulier aux normes en vigueur dans les différents pays ;

– attention aux codes mnémoniques : bien que la tendance naturelle soit d'attribuer des codes « parlants », il est préférable d'utiliser des codes non significatifs, car cela facilite l'évolution de la table contenant les codes ; imaginons que l'on souhaite codifier les personnes d'un service. L'utilisation du trigramme 1re lettre du prénom + deux premières lettres du nom serait certainement appréciée… dans un premier temps :

- Pierre Dupont : PDU,
- Christophe Brasseur : CBR,
- Patrick Laurent : PLA,
- Nathalie Durand : NDU ;

cette codification n'est pas évolutive car si une personne supplémentaire rejoint le service, par exemple Catherine Briand, comment la distinguer de Christophe Brasseur ? mieux vaut attribuer des codes séquentiels non significatifs : 001 = Pierre Dupont, 002 = Christophe Brasseur, etc. ; les codes ne constituent qu'une clé, c'est le libellé associé qui est important ;

– définir des règles claires et simples pour les descriptions ou libellés ; par exemple, pour une base de données articles, la description doit être structurée de façon homogène ; les troncatures et autres abréviations sont à bannir, car susceptibles d'interprétations diverses ;

– prendre en compte les aspects multilangues : dans les grands groupes, les projets informatiques sont de plus en plus internationaux ; la codification des données doit donc prendre en compte cette contrainte ; si certaines tables permettent de gérer des entrées dans différentes langues pour une même clé, d'autres non ; dans ce dernier cas, la codification doit être reconnue par tous les pays utilisateurs ; si une langue de référence doit être choisie, ce sera généralement l'anglais qui sera adopté.

Prenons l'exemple d'un fichier articles dont certaines données sont gérées en plusieurs langues, tableau 7.1.

Dans cet exemple, la désignation est gérée dans plusieurs langues. L'utilisateur voit à l'écran la description de l'article correspondant à la langue de connexion à l'application. Cet

exemple montre aussi que les unités sont traduites : ici le code L est le même (il pourrait être différent), mais le libellé associé est traduit.

Code article	Code langue	Désignation	Unité de stockage	
12312	FR	EAU OXYGENEE 95 % VRAC	L	Litre
12312	EN	OXYGENATED WATER 95% BULK	L	Liter
75123	FR	FIBRE DE VERRE	M	Mètre
75123	EN	GLASSFIBER	M	Meter

Tableau 7.1. *Exemple de données articles gérées en plusieurs langues*

Il faut toutefois faire attention aux tables multilangues et ne pas oublier de saisir plusieurs entrées lors de chaque création. Par exemple, il est nécessaire de saisir la description dans les différentes langues gérées, chaque fois qu'un nouvel article est créé.

7.3. La documentation des données

Il est recommandé de documenter avec soin les données gérées au sein de l'entreprise. Chaque application fait l'objet d'une documentation, qui inclut théoriquement un chapitre sur le modèle de données et la définition des attributs.

En pratique, on constate que ce n'est pas systématique. Il faut donc veiller à maintenir un document recensant les données gérées et décrivant leur utilisation et les principales règles de gestion. Cela présente un intérêt en particulier pour les ERP qui utilisent un grand nombre d'informations partagées par plusieurs domaines fonctionnels.

Ce document peut être organisé par catégorie de données, par exemple :

– données de référence : unités, devises, modes de paiement…

– données de structure organisationnelles : société, site, magasin, organisation commerciale…

– données de base : article, client, fournisseur, nomenclature…

– données dynamiques : commande client, commande d'achat, stocks…

Pour chaque donnée recensée, nous recommandons d'indiquer la définition, l'utilisation qui en est faite dans l'entreprise, les grandes règles de gestion, le principe de codification retenu, les contraintes particulières.

L'intérêt d'un tel document, véritable dictionnaire des données de l'entreprise, est de centraliser la connaissance du patrimoine informationnel et des règles de gestion. Il doit être consultable par les équipes informatiques comme par les utilisateurs. Aussi, ce référentiel est très utile pour l'évaluation de la qualité des données, car il permet de comparer la cible à la réalité.

7.4. L'administration des données

Lors de la mise en place d'une nouvelle application, il est essentiel de définir les procédures de gestion des données, en identifiant qui fait quoi.

Si cet aspect fait partie intégrante des méthodes de conception d'applications spécifiques de type Merise, il convient de ne pas le négliger lors de la mise en place de progiciels.

Il y a deux aspects à prendre en compte et à ne pas confondre :

– l'architecture logique de l'application,

– l'organisation de la gestion des données.

7.4.1. *L'architecture logique*

L'architecture détermine les systèmes logiques nécessaires au bon fonctionnement des applications. Ainsi, il peut y avoir un intérêt à mettre en place un système de référence des données de base, si une même application est implantée sur différents systèmes partageant les mêmes données.

Supposons qu'une application soit implantée sur trois systèmes géographiquement situés aux Etats-Unis, en Europe et en Asie (voir figure 7.1). Si les trois applications échangent des données communes entre elles (clients, articles, fournisseurs par exemple), il y a un intérêt à centraliser ces données au sein d'un référentiel et de les diffuser vers les systèmes qui en ont l'utilité. Avec ce principe, un client, un article ou un fournisseur n'est codifié qu'une seule fois.

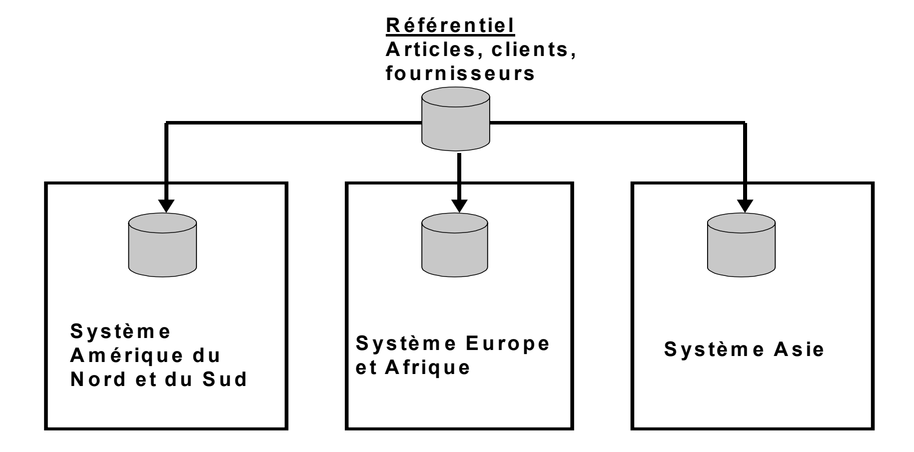


Figure 7.1. *Exemple d'architecture avec référentiel commun*

La mise en place d'un référentiel de données facilite les échanges entre les trois systèmes, puisque les données partagées ont les mêmes codes. Aussi, cela permet d'établir des reporting plus facilement, comme par exemple le chiffre d'affaires mondial par client ou encore par article. Toutes les données ne doivent cependant pas être maintenues de façon centrale. Il peut

être judicieux par exemple de décider de maintenir de façon centrale les données générales et communes de l'article (code, désignation, unité de stockage) et de maintenir localement les données propres à chaque région (données magasins et comptables par exemple).

Le fait d'avoir un référentiel pour certaines données ne signifie pas qu'une seule équipe maintient les données dans le système central. Il faut bien distinguer l'architecture des systèmes, de l'organisation mise en place pour gérer les données utilisées. Dans notre exemple, on peut très bien imaginer que le référentiel soit accessible aux trois équipes locales, même si le référentiel se trouve physiquement à New York.

La figure 7.1 est un exemple d'architecture, car d'autres possibilités existent. Un système mondial unique est une alternative à plusieurs systèmes communiquant entre eux. Les progrès dans les capacités des machines permettent aujourd'hui d'envisager ce type d'architecture.

7.4.2. *L'organisation de la gestion des données*

Il s'agit d'identifier les personnes qui vont maintenir les données et de définir les procédures de gestion associées. Ce point paraît évident, mais il est hélas fréquent de constater que les procédures de gestion des données ne sont pas clairement définies lors du démarrage d'un nouveau système.

Nous recommandons de déléguer la gestion des données propres à un domaine particulier aux utilisateurs naturels de ces données. Par exemple, les commandes d'approvisionnement seront naturellement saisies par les approvisionneurs. Les commandes de vente seront saisies par les assistantes commerciales.

Les données dynamiques ne posent en général pas trop de problème. La difficulté réside plutôt du côté des données de base

telles que les articles, car différentes fonctions sont concernées : logistique, stocks, comptabilité, contrôle de gestion. Comment créer alors un nouvel article ?

Pour répondre à ce type de questions, il y a besoin de définir la procédure la plus adaptée, en faisant participer les différents acteurs impliqués dans la maintenance de la donnée. Généralement, la mise en place d'un circuit faisant intervenir différents acteurs est nécessaire et il est important de le formaliser par un schéma aussi clair que possible. Les méthodes traditionnelles de conception proposent des schémas et diagrammes particulièrement adaptés à la description de l'enchaînement des tâches et du « qui fait quoi ».

Prenons l'exemple de la création d'un article dans un ERP : un grand groupe industriel comporte plusieurs entreprises. La demande de création d'une nouvelle matière première par l'une des entreprises est validée par la structure centrale des achats du groupe. Aussi, l'initialisation dans le système est également centralisée au sein d'une cellule administration des données. Cette cellule veille en particulier à ne pas introduire de doublons dans le système. Une fois les informations générales créées, les différentes équipes opérationnelles de l'entreprise complètent les autres informations (approvisionnement, stocks, production, comptabilité), chacune de ces équipes ayant accès a une vue dédiée dans le progiciel.

Ce type de schémas (figure 7.2) doit être utilisé et communiqué accompagné d'explications.

La centralisation de certaines données doit être étudiée avec le plus grand soin. Il n'y a pas de règle générale permettant d'affirmer qu'il faut ou non centraliser la gestion de telle ou telle donnée.

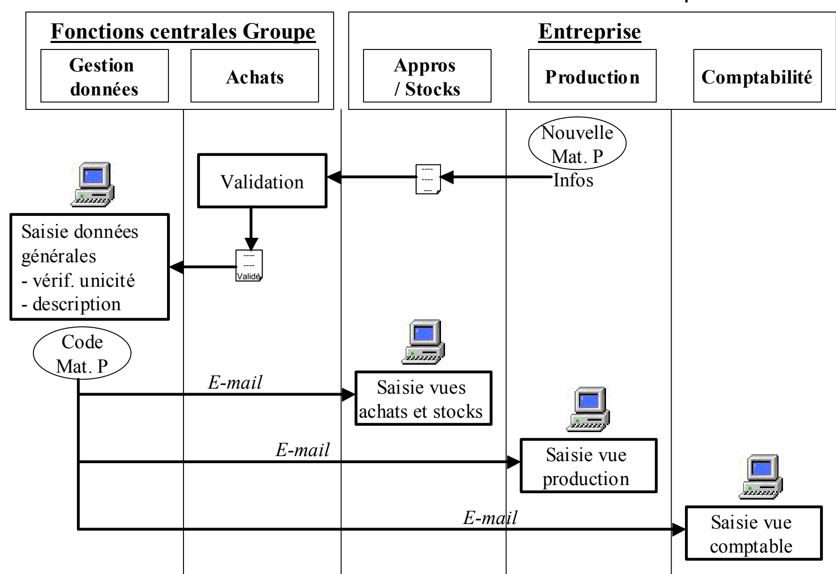


Figure 7.2. *Exemple de procédure de création d'un article*

On peut en revanche déconseiller la centralisation de gestion des données nécessitant une connaissance particulière. Dans l'exemple précédent, la centralisation des vues MRP ou magasin n'aurait pas de sens. C'est aux utilisateurs de ces vues de les gérer. La centralisation des vues générales peut en revanche présenter un intérêt, puisque ces vues sont partagées par différents types d'utilisateurs et concernent tous les domaines. Ainsi, en se concentrant sur l'initialisation, une équipe centrale portera davantage attention aux normes ou aux doublons.

D'autres façons de faire pourraient cependant tout aussi bien convenir (par exemple décider que ce sont les approvisionneurs qui initialisent la matière première), à condition que des règles claires de normalisation, de vérification de non-existence, etc. soient bien établies et que des contrôles réguliers soient effectués.

Mode	Avantages	Inconvénients
Centralisé	▪ Respect des normes ▪ Evite les doublons ▪ Allège le travail des opérationnels	▪ Réactivité (un maillon de plus dans la chaîne) ▪ Coût (car besoin d'une équipe dédiée)
Décentralisé	▪ Réactivité ▪ Autonomie ▪ Meilleure appropriation de la donnée par l'utilisateur	▪ Risque de dériver par rapport à un langage commun ▪ Risque de doublons accru ▪ A terme, pollution de la base de données (doublons…)

Tableau 7.2. *Avantages et inconvénients d'une gestion centralisée des données*

Le tableau 7.2 présente les avantages et inconvénients de chacun des deux modes d'administration des données.

7.5. Le nettoyage des bases de données (*clean-up*)

7.5.1. *Définition*

Le nettoyage (*clean-up*), régulier ou occasionnel des bases de données est une opération assez courante. Il consiste à éliminer les données obsolètes ou sans objet et à corriger les erreurs d'informations stockées dans la base de données. Cette opération procure un résultat concret dans un temps relativement court.

Plutôt qu'une opération technique, le clean-up est d'abord une démarche répondant à des besoins opérationnels de recherche d'efficacité. A titre d'exemple, il peut s'agir de supprimer des commandes créées sans objet. Ce peut être aussi une opération visant à éliminer les doublons dans la base clients ou articles. Ou encore une opération de correction des

enregistrements de la base de données fournisseurs (harmonisation des adresses par exemple).

Bref, il s'agit d'actions concourant à :

– mettre la base de données en cohérence avec les règles de gestion ;

– supprimer les données inutiles.

L'opération est facilitée par un certain nombre d'outils automatisant la détection d'erreurs. Bien sûr, la détection d'erreurs est une chose, la correction en est une autre. La correction consiste à modifier chaque donnée incorrecte pour la mettre en phase avec les règles de gestion. C'est un travail qui ne peut pas toujours être automatisé. Les deux activités (détection et correction) sont menées habituellement en séquence et permettent de repartir sur une base saine dans un délai raisonnable.

Il y a souvent une grande satisfaction de l'entreprise et des managers après des opérations de nettoyage de bases de données.

7.5.2. *Dans quels cas faut-il nettoyer les données ?*

Le nettoyage peut être effectué ponctuellement ou au contraire régulièrement. Mais il y a des cas où il n'est pas forcément applicable.

Prenons l'exemple d'une société de Bourse relativement importante qui enregistre 10 000 ordres par jour. La base de données des ordres s'enrichit donc de 10 000 enregistrements par jour. Supposons que 3 % des enregistrements contiennent une erreur. Cela représente donc 300 enregistrements erronés chaque jour. Si cette société choisit de procéder à des nettoyages de façon régulière, cela signifie qu'il faut corriger 300 enregistrements chaque jour, ce qui représente évidemment une charge de travail importante.

Il ne s'agit-là que d'un exemple avec des volumes relativement modestes. Que penser de certaines activités qui nécessitent l'enregistrement de millions de données chaque jour, comme par exemple les organismes d'assurance maladie ou encore certains opérateurs de télécommunications ?

On peut donc conclure que le nettoyage des bases de données est difficilement applicable aux organisations qui génèrent de grandes quantités de données. Le nettoyage de données peut en revanche être intéressant pour des données stables, ayant un taux de croissance assez faible.

Lors de la migration de données d'une application vers une autre, le nettoyage avant intégration des données constitue le second cas intéressant. En effet, cela permet de démarrer une nouvelle application avec des données saines et de bénéficier pleinement de l'apport fonctionnel du nouveau logiciel.

Un dernier point doit enfin être pris en considération : il s'agit du périmètre des données à prendre en compte. En effet, il est fréquent de constater qu'un nombre élevé de données présentes dans les bases de données ne servent à rien. Ainsi, certaines entreprises reconnaissent que près de la moitié de leurs états et « reports » ne sont jamais utilisés.

Il est donc conseillé de se consacrer aux données vitales et stratégiques de l'entreprise et d'ignorer sans états d'âme les données non utilisées ou sans importance. En conclusion, nous pouvons retenir les principes suivants :

– procéder à des nettoyages :
  - lors de l'initialisation d'une application,
  - ou pour des données stables à faible taux de croissance ;
– se focaliser sur les données utiles et stratégiques ;
– ignorer les données sans importance.

7.5.3. *L'alternative au* clean-up

Comme le montre le paragraphe précédent, le nettoyage des bases de données ne peut pas s'appliquer à l'ensemble des données, en raison des inconvénients suivants :

– le coût : pour être efficace, cette méthode doit être appliquée régulièrement pour ne pas dire en permanence ; car à peine nettoyée, la base de données s'enrichit de nouveaux enregistrements qui ne sont pas nécessairement corrects ; un mois après un nettoyage, un nombre non négligeable de nouvelles données incorrectes est déjà là ; le nettoyage permanent coûte donc très cher ;

– l'efficacité du nettoyage : ne nous leurrons pas, beaucoup d'erreurs passent au travers de l'étape de détection ; les outils de détection identifient les erreurs évidentes mais laissent de côté les erreurs plus subtiles portant par exemple sur des données obéissant à des règles de gestion complexes ; quant à la détection manuelle, donc humaine, elle est évidemment d'une efficacité limitée, compte tenu du volume important des données à vérifier.

Clairement, le nettoyage des données seul ne suffit pas à améliorer durablement la qualité des données. Il est donc plus judicieux d'adopter *une démarche préventive* visant à limiter les anomalies à la source. Cette approche consiste à identifier les causes d'erreurs et à les éliminer. Par exemple, si la cause de non-qualité est liée à la saisie, des solutions de type formation, renforcement des contrôles à la saisie ou mise en œuvre d'un système de lecture code à barres seront privilégiés. Cette méthode est largement décrite aux chapitres 5 et 6. Elle privilégie l'analyse de l'ensemble de la chaîne d'information et non uniquement les données. Lorsque cette approche est adoptée et réalisée avec rigueur, les résultats sont spectaculaires. Les taux d'erreurs peuvent facilement être divisés par 10.

L'approche préventive n'est pas naturelle, car la tendance générale est de résoudre les problèmes immédiats et de reporter

à plus tard les améliorations à plus long terme. C'est un tort, car le bénéfice obtenu est sans commune mesure avec les autres approches.

Comparons les trois grandes approches possibles :

– approche 1 : gestion des problèmes au fil de l'eau, sans intervention, ni sur la chaîne d'information, ni sur la base de données ;

– approche 2 : clean-up ;

– approche 3 : démarche préventive.

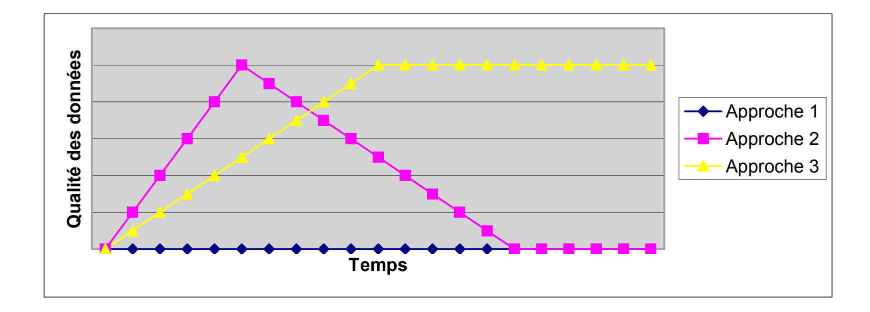


Figure 7.3. *Impacts sur la qualité des données selon la démarche*

Le schéma de la figure 7.3 illustre l'impact de chacune des trois grandes approches décrites précédemment sur la qualité des données :

– l'option 1 n'améliore aucunement la qualité des données ;

– l'option 2 (nettoyage) engendre assez rapidement un niveau de qualité satisfaisant, mais celui-ci retombe progressivement au niveau de départ ; cette approche répétée constamment permettrait de maintenir un niveau correct, mais coûterait très cher ;

– l'option 3 (préventive) améliore progressivement la qualité des données et maintient celle-ci dans le temps à un niveau élevé et stable. L'option préventive 3 est donc globalement préférable à l'option corrective 2. Rappelons cependant que l'option 2 est intéressante lors de la migration des données d'un système vers un autre.

7.5.4. *Le cas des doublons (ou* **n-uplets***)*

Les doublons sont un véritable problème. Ils correspondent à des enregistrements multiples d'une même réalité. Par exemple, un même article ou un même client est enregistré plusieurs fois dans la base de données. On imagine aisément les soucis de gestion que cette situation engendre (voir paragraphe 4.2.1).

Ce problème n'est pas simple à traiter, notamment au sein des organisations qui gèrent de grandes quantités d'informations. Deux types d'actions peuvent néanmoins être menées :

– la détection et l'élimination des doublons dans la base de données existante,

– le contrôle des doublons lors de la saisie d'un nouvel enregistrement.

7.5.4.1. *Détection des doublons dans la base de données*

Généralement le volume des données enregistrées (plusieurs milliers d'enregistrements) est tel qu'un contrôle manuel est difficilement envisageable.

L'utilisation d'outils de détection de doublons facilite la tâche de recherche. Ces outils sont basés pour la plupart sur la phonétisation. Ils permettent la recherche de doublons ou n-uplets engendrés par les différents aléas affectant la saisie et la collecte de l'information : saisie incomplète, variantes orthographiques des raisons sociales ou désignations, substitution des raisons sociales par leurs acronymes, inversion des mots ou du contenu des champs, etc. La recherche se fait sur un ou plusieurs champs (nom, adresse…). L'utilisation de ces logiciels s'applique à la recherche de doublons articles, clients, fournisseurs…

Il existe également des solutions de dédoublonnage des adresses, qui s'appuient sur des fichiers d'organismes externes (La Poste…). Ces solutions sont évidemment adaptées à la

recherche de doublons au sein de fichiers présentant des adresses comme les fichiers clients ou fournisseurs.

Si l'on ne souhaite pas acquérir d'outil, il est aussi possible de sous-traiter ce type d'opérations à des sociétés spécialisées.

Avant de choisir telle ou telle solution, il est recommandé de définir la problématique actuelle, mais aussi future et de chiffrer le coût des différentes alternatives.

7.5.4.2. *Contrôle des doublons lors de la saisie d'un nouvel enregistrement*

Là encore, il n'y a pas de solution miracle. On peut soit gérer le contrôle manuellement, soit avoir recours à des outils :

– contrôle manuel : il s'agit de vérifier que l'enregistrement que l'on souhaite créer n'existe pas déjà ; pour les données sans adresse, comme par exemple les articles, une recherche sur la désignation doit systématiquement être effectuée avant toute nouvelle création ; pour les données avec adresses (clients, fournisseurs), une recherche sur la raison sociale et l'adresse doit être effectuée avant de saisir un nouvel enregistrement ; cette méthode a évidemment des limites, mais si les utilisateurs possèdent le réflexe d'effectuer une recherche avant toute création, le nombre de cas de doublons sera limité ;

– contrôle automatisé : certaines données tel que le code Siret peuvent être très utiles pour ne pas générer de doublons ; un contrôle automatisé sur ce type de champ peut alors être mis en place ; l'exemple du code Siret est néanmoins limité à l'Hexagone ; si l'on souhaite étendre ce principe à une base de données mondiale, il existe des sociétés (D&B…) qui codifient la plupart des entités juridiques mondiales en attribuant un code distinct par adresse géographique ; il peut être intéressant de gérer ce type de code dans la base de données pour distinguer chaque enregistrement et mettre en place un contrôle à la saisie ; aussi, certains outils basés sur la phonétisation peuvent être utilisés en temps réel.

En conclusion, le problème des doublons n'est pas simple. Une gestion soit manuelle et volontaire, soit basée sur l'utilisation d'outils, diminue néanmoins les risques.

7.6. La surveillance des données

La surveillance des données doit être constante. Les méthodes développées dans les chapitres précédents permettent de mesurer le niveau de qualité de l'information et de prévenir toute dégradation.

La vigilance vis-à-vis des doublons doit être permanente. Les bases de données clients, fournisseurs et articles doivent ainsi être régulièrement observées, en s'aidant si besoin d'outils dédiés.

Aussi, une attention particulière doit être portée aux flux complexes traversant plusieurs domaines fonctionnels. C'est le cas notamment d'applications différentes reliées entre elles par des interfaces. Une modification mineure dans une application amont peut en effet engendrer un désastre dans une application avale (voir paragraphe 4.3.3).

Les architectures distribuées sont de plus en plus répandues. Si elles apportent davantage de souplesse et d'autonomie, elles engendrent aussi des risques en termes d'intégrité des données échangées. Il est donc recommandé d'utiliser des outils de monitoring pour suivre les échanges de données et analyser les erreurs.

7.7. La sécurité des données

7.7.1. *Sécurité de la saisie des données*

Lorsque l'information est saisie manuellement *via* un clavier, la sécurité de la saisie est généralement assurée par des contrôles programmés. Les contrôles ont pour but de vérifier la validité et

la cohérence des données saisies par l'utilisateur (exemple : contrôle du code TVA). Il ne faut pas confondre ce type de contrôle avec les contrôles d'accès aux données (autorisations).

Tous les contrôles élémentaires doivent être intégrés au niveau de la saisie. Cela réduit considérablement le temps de traitement de l'erreur, puisque seule une correction permet de sortir de l'écran. Il est à noter que les progiciels intègrent en standard un grand nombre de contrôles, y compris des contrôles croisés entre zones.

Les messages d'erreurs rapatriés par le système doivent être explicites pour aider l'utilisateur à corriger les erreurs de saisie. Aussi, l'analyse et le traitement des messages d'erreurs doivent être intégrés à la formation des utilisateurs. Ceux-ci doivent en particulier être capables de différencier une erreur système d'une erreur de saisie.

Une autre possibilité limitant les erreurs de saisie de façon spectaculaire est l'utilisation du *code à barres*. Le code à barres est aujourd'hui la solution technique la plus utilisée pour acquérir automatiquement une information. Parmi ses qualités, citons :

– la fiabilité et la rapidité de lecture, avec des taux d'erreurs de lecture allant d'une erreur sur 20 000 à 2 000 000 de codes lus (selon le type de codes lus) ;

– la redondance de l'information sur toute la hauteur du code, qui garantit la justesse de l'information, même si une partie du code est illisible suite à des défauts d'impression, des frottements ou encore des déchirements légers ;

– la facilité et le faible coût d'impression du code à barres : de nombreuses technologies, supports et consommables d'impression sont disponibles. Le code à barres peut aisément accompagner d'autres informations nécessaires à un étiquetage et voit, de fait, son coût de revient diminuer.

La technologie code à barres ne s'applique pas à tous les domaines, mais est largement utilisée :

– dans la distribution : elle permet d'entrer rapidement un article en stock, de connaître l'origine du produit, de faciliter le réapprovisionnement, d'automatiser la sortie des produits et d'obtenir un inventaire simplifié permanent et bien sûr d'optimiser les temps de traitement aux caisses ;

– dans l'industrie : elle permet dans ce cas de connaître l'origine des matières premières, d'identifier sa production et d'enregistrer plus facilement les expéditions transmises à ses clients.

7.7.2. *Sécurité du traitement des données*

Un programme peut révéler des erreurs après quelques temps d'utilisation. Il peut s'agir d'erreurs de conception ou de maintenance du programme, par exemple suite à une modification de version.

Les programmes de traitement doivent donc être testés en réalisant des jeux de tests appropriés. Pour les traitements en chaîne, des points de contrôle (états, rapports) doivent être mis en place, pour pouvoir reconstituer le cas échéant la trace des opérations réalisées.

7.7.3. *Sécurité des accès à l'application*

On distingue plusieurs niveaux de contrôle d'accès :

– le contrôle des habilitations : il s'agit, d'une part, d'interdire l'accès au système d'information de personnes non habilitées, d'autre part, d'être capable d'identifier où, quand et comment des personnes habilitées ont accédé à la machine et à la base de données ;

– le contrôle des autorisations : il s'agit de permettre l'accès de personnes habilitées à certaines fonctionnalités et pas

d'autres ; un comptable par exemple, pourra visualiser des commandes d'achats, mais ne pourra pas en créer ; cette dernière fonctionnalité sera réservée aux approvisionneurs et aux acheteurs.

La définition des *rôles* et des autorisations associées doit être étudiée avec la plus grande attention. En fonction de son métier, chaque utilisateur doit être affecté à un ou plusieurs rôles, lui donnant ainsi le droit d'accéder à telle ou telle donnée (consultation, modification ou création).

7.8. L'auditabilité des données

Le contrôle fiscal des comptabilités informatisées s'inscrit dans un cadre légal qui comprend un certain nombre de dispositions. L'Administration fiscale dispose de plusieurs Brigades de vérification des comptabilités informatisées (BVCI) regroupant des agents formés spécifiquement à cet effet. Les contrôles peuvent porter sur l'ensemble des informations, données et traitements qui concourent directement ou indirectement à la formation du résultat comptable et fiscal ainsi qu'à l'établissement des déclarations obligatoires.

Les entreprises doivent donc conserver :

– la trace des données et opérations depuis l'origine jusqu'à la comptabilité (saisie),

– les historiques des mouvements (commandes, livraisons, mouvements de stocks, factures, etc.),

– les fichiers de référence (plan comptable, tarifs, contrats, fichiers clients, fournisseurs, etc.).

Les délais de conservation des documents d'entreprise sont déterminés soit par la loi, soit en fonction des délais de prescription applicables et des périodes pendant lesquelles les administrations peuvent effectuer des contrôles. Les paragraphes

suivants précisent les différents délais qui peuvent être opposés aux entreprises.

7.8.1. *Les documents à conserver pendant trente ans*

Les entreprises sont tenues de garder les contrats et les documents relatifs à leurs titres de propriété pendant trente ans ou plus, s'il s'agit des contrats à exécution échelonnée. Voici quelques exemples :

– statuts et actes de société,
– rapports des Commissaires aux comptes,
– rapports du Conseil d'administration,
– comptes sociaux,
– comptes d'exploitation,
– comptes de pertes et profits.

7.8.2. *Les documents à conserver pendant dix ans*

Les documents et les pièces justificatives afférentes aux éléments de la comptabilité des commerçants doivent être conservés dix ans. Par exemple :

– documents comptables :
  - livre journal,
  - grand livre,
  - bilans, comptes de résultat,
  - balance générale annuelle,
  - journaux et comptes auxiliaires,
  - factures client,
  - factures fournisseurs,
  - tout document relatif aux impôts, taxes et droits ;
– documents commerciaux :
  - correspondances commerciales,
  - commandes clients,
  - commandes fournisseurs.

7.8.3. *Les documents à conserver pendant cinq ans*

Doivent être conservés par l'entreprise pour pouvoir être présentés à l'administration fiscale, les documents relatifs à son résultat fiscal des trois ou cinq dernières années, en cas de déficit antérieur imputé sur une de ces trois ou cinq dernières années ou en cas d'amortissements différés. Par exemple :

– registres et états du personnel ;

– toutes pièces comptables relatives aux traitements et salaires : cartes de pointage, décomptes de commissions …

– reçus pour solde de tout compte ;

– doubles des certificats de travail ;

– doubles des bordereaux de paiement des cotisations sociales ;

– journaux et récapitulatifs des charges sociales ;

– fiches fiscales par salarié.

7.8.4. *Les documents à conserver pendant trois ans*

C'est le délai général de reprise de l'administration fiscale et le délai pendant lequel elle peut se faire communiquer :

– la documentation comptable (livres, registres, etc.) établie sur support informatique,

– la documentation relative aux analyses, à la programmation et à l'exécution des traitements informatiques.

Tous documents de transport intéressant les douanes doivent être conservés trois ans. Ces délais ne signifient pas que tous les documents listés ci-dessus doivent être consultables « en ligne ». Simplement les documents doivent être disponibles et accessibles sous une forme ou une autre. Des techniques d'archivage permettent de migrer les informations sur des supports autres que la base de données courante.

7.9. Conclusion

Un certain nombre de bonnes pratiques (compréhension des besoins, codification, procédures…) sont nécessaires à la qualité des données. Ces principes ne concernent pas uniquement la phase de conception du système d'information ; ils doivent aussi être appliqués après démarrage des applications et donner lieu à une surveillance permanente.

Quatrième partie

La reprise des données

CHAPITRE 8

Stratégie de reprise des données

8.1. Définition de la reprise des données

La reprise des données consiste à migrer des données d'une ancienne application informatique vers une nouvelle. La reprise des données doit en particulier tenir compte :

– des aspects légaux : la migration des données ne doit pas engendrer d'anomalies comptables ; c'est la raison pour laquelle la plupart des projets à caractère financier prévoient le démarrage de la nouvelle application après une date de clôture qui gèle les soldes et interdit l'accès aux périodes clôturées dans l'ancien système ;

– des contraintes métier : il faut éviter des perturbations de l'activité et ne pas gêner l'exécution des tâches opérationnelles ; par exemple les expéditions aux clients ne doivent pas être retardées sous prétexte que l'application change.

Comme schématisée par la figure 8.1, la problématique est donc bien d'assurer une continuité dans un processus en rupture : le système d'information change à une date donnée, tandis que les flux physiques et informationnels doivent continuer à vivre leur vie.

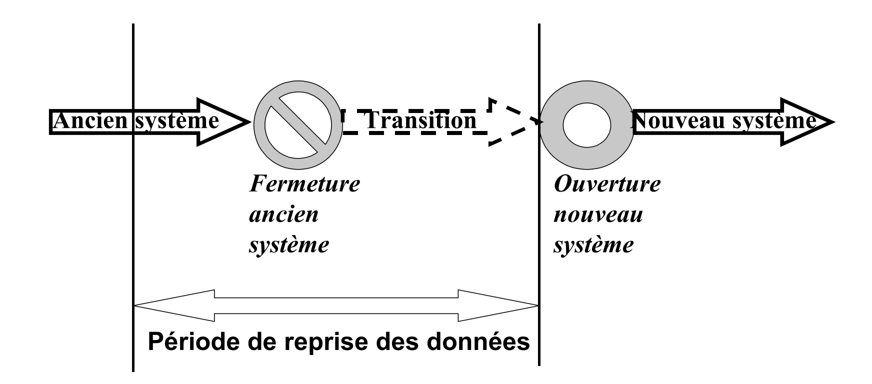


Figure 8.1. *Transition entre ancien et nouveau système*

La période de transition, c'est-à-dire la période comprise entre la fermeture de l'ancien système et l'ouverture du nouveau système doit être la plus courte possible. Pour ne pas pénaliser le travail opérationnel, la transition s'effectue généralement le week-end, mais peut comporter un ou deux jours supplémentaires. Puisqu'il n'y a pas de système informatique disponible pendant cette période, il y naturellement besoin de préparer des documents papiers à l'avance afin de tracer toute activité qui sera régularisée dans le nouveau système. Par exemple, des tableaux permettant de tracer les mouvements de stocks doivent être édités.

Les procédures de reprise des données peuvent être complexes, suivant le volume et la difficulté des conversions de données à effectuer. Globalement on distingue trois étapes majeures dans le processus de reprise :

– l'extraction des données de l'ancien système ;

– la conversion du contenu de l'extraction pour l'adapter au futur système ;

– le chargement ou la saisie des données dans le nouveau système.

Ces trois étapes sont positionnées dans le temps en fonction du type de données.

8.2. Différents types de données et mode de reprise

La migration des données nécessite d'analyser et de lister les données à reprendre dans le nouveau système. Il est également nécessaire de définir la méthode de reprise (automatique ou manuelle) et la date de reprise.

Type	Donnée/objet de données	Méthode	Date de reprise
Données de référence	Unités de mesure	Manuel	M - 4
	Devises	Manuel	M - 4
	Modes de paiement	Manuel	M - 4
	Conditions de paiement	Manuel	M - 4
Données structurantes	Sociétés	Manuel	M - 3
	Sites	Manuel	M - 3
	Magasins	Manuel	M - 3
	Emplacements	Manuel	M - 3
	Structures commerciales	Manuel	M - 3
	Structures d'achats	Manuel	M - 3
	Centres de coûts	Manuel	M - 3
Données de base	Clients	Automatique	M - 2 à J
	Articles	Automatique	M - 2 à J
	Fournisseurs	Automatique	M - 2 à J
	Nomenclatures de production	Automatique	M - 2 à J
Données dynamiques	Commandes clients non soldées	Manuel	J – 10 à J
	Commandes fournisseurs non soldées	Manuel	J – 10 à J
	Stocks	Automatique	J – 2 à J
	En-cours comptables (clients, fournisseurs, banques, factures non parvenues)	Automatique	J – 2 à J
	Prix des articles	Automatique	J – 5 à J

Tableau 8.1. *Exemple de mode et calendrier de reprise selon le type de données*

Les données stables sont généralement migrées bien avant le démarrage de la nouvelle application, tandis que les données dynamiques, comme les stocks ou les en-cours comptables, sont migrées au dernier moment. La méthode de reprise dépend du volume : si celui-ci est important, on privilégie une reprise automatique ; dans le cas contraire, une reprise manuelle est plus adaptée. Le tableau 8.1 est un exemple de classification des données avec pour chacune d'elles, la méthode de reprise (automatique par programme ou saisie manuelle).

Il ne s'agit-là que d'un exemple, qui peut bien entendu varier d'un contexte à un autre. Notons que le terme de données est parfois un raccourci : il s'agit en fait d' « objets » de données. Par exemple, « client » est un objet de données comprenant un ensemble de données élémentaires (attributs ou zones).

Le travail d'inventaire des objets de données à migrer doit être anticipé au maximum, afin de développer et tester à temps les outils d'extraction et de chargement.

Aussi, une attention particulière doit être portée à la gestion des nouvelles données créées ou modifiées après extraction. Par exemple, une fois que les données « clients » sont extraites pour être chargées dans le nouveau système, il se peut que de nouveaux clients soient créés dans l'ancien système. Il est donc nécessaire de piéger toute nouvelle création ou modification après extraction et créer un lot complémentaire de chargement dans le nouveau système.

8.3. Méthodologie de reprise des données

La migration des données nécessite un certain nombre de tâches préparatoires permettant de réaliser la reprise le moment venu.

8.3.1. ***La préparation de la reprise***

Un travail préparatoire comprend les tâches essentielles décrites ci-après.

8.3.1.1. *Nettoyage des données*

Le passage à un nouveau système est l'occasion de nettoyer les données : élimination des données obsolètes, correction de données, dédoublonnage… Le nettoyage peut s'effectuer directement dans le système source, mais il est souvent plus simple de le réaliser dans les fichiers d'extraction (sous Excel par exemple). Bref, il s'agit de mener une opération de clean-up comme recommandé à la section 7.5. Aussi les outils de dédoublonnage peuvent être particulièrement utiles à cette étape.

Outre le nettoyage des données, il convient de s'assurer que les données sont en phase avec la réalité. Cela est particulièrement important pour certaines données dynamiques comme par exemple les stocks. Dans ce cas précis, un inventaire physique doit être effectué quelques semaines avant la migration, pour que les données du système source soient justes au moment de la reprise.

8.3.1.2. *Spécification des règles de gestion*

La première tâche à effectuer est d'identifier les données à reprendre. L'analyse de l'application et des processus cibles permet de déduire la liste des objets de données nécessaires au fonctionnement de la nouvelle application.

Lorsque les objets de données de l'application cible sont identifiés, il faut alors définir précisément les règles de gestion et de codification de chacune des zones de ces objets. Si la définition est floue, la reprise le sera également et conduira à des défauts de qualité.

Chaque donnée élémentaire nécessaire au système cible est spécifiée : liste de valeurs possibles, caractère obligatoire ou facultatif, signification, contraintes particulières.

Pratiquement, l'ensemble des règles de gestion doivent être décrites dans un ou plusieurs documents de spécifications rédigés soigneusement. Peu importe la forme (texte ou tableau Excel), l'important est que les règles de gestion soient clairement décrites.

8.3.1.3. *Définition des règles de conversion (*mapping*)*

A partir de chaque donnée ou objet de données nécessaire au système cible, il faut identifier l'équivalent dans l'ancien système (système source) et décrire les règles de conversion.

Par exemple la valeur « 03 » de tel mode de paiement de l'ancien système est désormais codifiée en « 005 » dans le nouveau. Les règles de conversion sont utiles à la préparation des structures de chargement. Il est fortement recommandé d'associer à ce travail les personnes du métier concerné.

Aussi, il est fréquent qu'une nouvelle application nécessite des données qui n'étaient pas gérées auparavant. Dans ce cas, il faut étudier la meilleure façon d'initialiser les données en question dans le nouveau système (saisie manuelle, programme). Cette étape doit être effectuée avec soin, car elle conditionne la qualité des données dans le système cible. En pratique, des tableaux de conversion doivent être documentés.

8.3.1.4. *Détermination du mode de reprise (automatique/manuel)*

Donnée par donnée, en tenant compte à la fois du volume et de la complexité des règles de migration, il s'agit de déterminer si la reprise s'effectue de façon automatique ou manuelle. Rappelons qu'il y a deux étapes à considérer : l'extraction et le chargement.

Dans certains cas, il peut être envisagé d'effectuer l'extraction automatiquement et le chargement manuellement. Cette étape nécessite d'être pragmatique et de vraiment chercher le moyen le plus simple et le plus sûr pour intégrer les données.

8.3.1.5. *Définition de la séquence de chargement (ou de saisie) dans le nouveau système*

La séquence de reprise des différents objets de données doit être précisée. Par exemple, avant de reprendre les données comptables d'un client, il faut d'abord créer les données générales (raison sociale, adresse, contacts, etc.). L'ordre des chargements automatiques est généralement identique à celui d'une saisie en temps réel et repose la plupart du temps sur des règles de bon sens.

8.3.1.6. *Développement des outils de conversion*

Lorsque les règles de conversion (*mapping*) sont établies, il reste à développer des outils de conversion permettant d'extraire, de transformer et de charger les données. Cela concerne bien entendu les données à reprendre de façon automatique. Il y a besoin à la fois d'outils d'extraction du système source et d'outils de chargement dans le système cible. La phase de transformation peut être manuelle ou automatisée, selon la complexité des règles de conversion. Une alternative au développement sur mesure de programmes de migration consiste à utiliser des outils dédiés aux conversions (voir section 8.4).

8.3.1.7. *Définition de la structure des fichiers de reprise*

En parallèle à l'élaboration des outils de reprise, la structure des fichiers permettant le chargement automatique dans la nouvelle application sont définis. Ces fichiers constituent le point d'entrée des programmes de chargement. Ces structures comprennent l'ensemble des zones de chaque objet de données à charger. Chacune des zones du fichier doit normalement être décrite dans le document des règles de gestion, comme l'indique la figure 8.2.

Structure du fichier article

PTRU-YRTE	PTRU-UJED	…	…
Désignation	Unité		

Règles de gestion de l'article

Nom technique zone	Nom	Format	Longueur	Valeurs possibles	Règles de gestion
PTRU-YRTE	Désignation	A	50	Texte libre	pa ta ti pa ta tat
PTRU-UJED	Unité	A	3	kg (kilogramme), l (litre)	pa ta ti pa ta tat
…					
…					

Figure 8.2. *Structure d'un fichier de reprise et règles de gestion*

8.3.2. *Planification et réalisation de la reprise*

Pour être réalisée dans les meilleures conditions, l'opération de bascule doit être organisée et planifiée à l'avance.

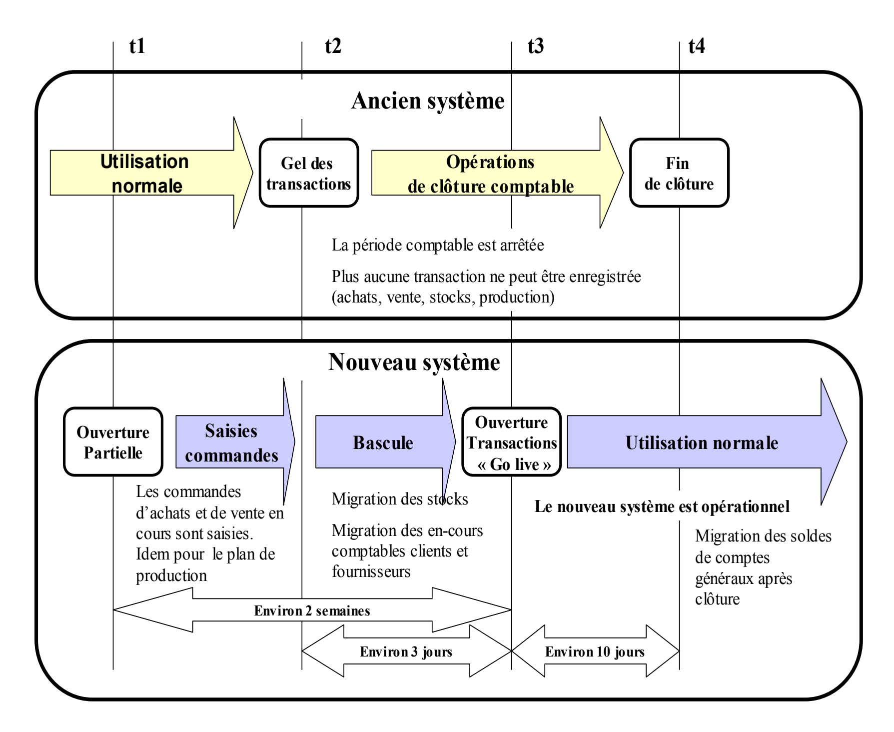


Figure 8.3. *Exemple de séquence des opérations de migration de données dynamiques*

Un planning détaillé décrivant l'ensemble des tâches et des ressources est établi et communiqué à l'ensemble des intervenants. Rappelons que si certaines données peuvent être migrées bien avant la bascule (comme les données de base), d'autres doivent être migrées au dernier moment dans un temps très court, comme les données dynamiques (voir tableau 8.1).

La figure 8.3 présente un exemple de la séquence des opérations de reprise de données dynamiques. Il s'agit du cas d'une migration relativement importante englobant les domaines comptables, production, vente, achats et stocks. Ce schéma met en évidence le chevauchement d'utilisation de l'ancien système et du nouveau système.Ce schéma fait apparaître les jalons suivants :

– l'ouverture partielle du nouveau système (t1) : environ deux semaines avant le démarrage, le nouveau système est ouvert à la saisie des données qui ont commencé leur vie dans l'ancien système et l'achèveront dans le nouveau ; c'est le cas par exemple des commandes d'achats saisies dans l'ancien système et dont la réception se fera après le gel de celui-ci ; dans cet exemple, la reprise est manuelle, mais pourrait être automatisée ;

– le gel des transactions dans l'ancien système (t2) : environ trois jours avant le démarrage, ce jalon marque la fin de la saisie de tout mouvement dans l'ancien système ; la date retenue correspond à une date de fin de mois comptable ; à partir de cette date, les opérations de migration des stocks, d'en-cours comptables peuvent commencer ; l'ancien système reste cependant ouvert aux opérations de clôture qui permettent d'ajuster les comptes généraux ; notons que entre t2 et t3, aucune transaction de type mouvement de stocks, déclaration de production, décaissement ou encaissement ne peut être effectuée ni dans l'ancien système, ni dans le nouveau ; il y a donc besoin de limiter ces mouvements et de s'organiser pour les tracer manuellement. Ces mouvements manuels seront ressaisis dans le nouveau système après migration des données dynamiques (régularisations) ;

– l'ouverture du nouveau système à l'ensemble des utilisateurs (t3) : lorsque la migration est terminée et que la nouvelle période comptable est ouverte dans le nouveau système, les utilisateurs peuvent désormais saisir de nouveaux mouvements dans la nouvelle application ; la première chose à faire est de régulariser tous les mouvements qui auront été effectués pendant la période de bascule ;

– la fin de la clôture comptable dans l'ancien système (t4) marque l'arrêt définitif de ce système ; les soldes des comptes généraux peuvent alors être repris manuellement ou automatiquement dans le nouveau système.

Cet exemple, décrit ici de façon synthétique, illustre bien l'importance de la synchronisation des différentes tâches liées à la reprise des données dynamiques. Pendant la bascule en particulier, l'ensemble des tâches à exécuter doit être orchestré et suivi de très près.

8.4. Les outils de conversion de données

Classiquement, la reprise des données s'appuie sur des programmes d'extraction du système source et des programmes de chargement dans le système cible. Entre extraction et chargement, la conversion des données est généralement effectuée de façon manuelle ou semi-automatique par manipulation de tableaux Excel. Les manipulations s'appuient sur les règles de conversion définies lors de l'étape de mapping.

Depuis plusieurs années, des outils de conversion permettent de mieux maîtriser la migration des données. Certains outils offrent des possibilités de nettoyage, de dé-doublonnage, de surveillance et d'alerte.

Le marché propose en particulier des outils « ETL » (*Extract, Transform and Load*) spécialisés dans la transformation des données et facilitant l'intégration des données. Les ETL sont

surtout utilisés pour alimenter des Datawarehouse, mais peuvent aussi être utilisés pour la reprise des données.

Un processus ETL se décompose en trois phases : l'extraction, la transformation et le chargement (voir figure 8.4).

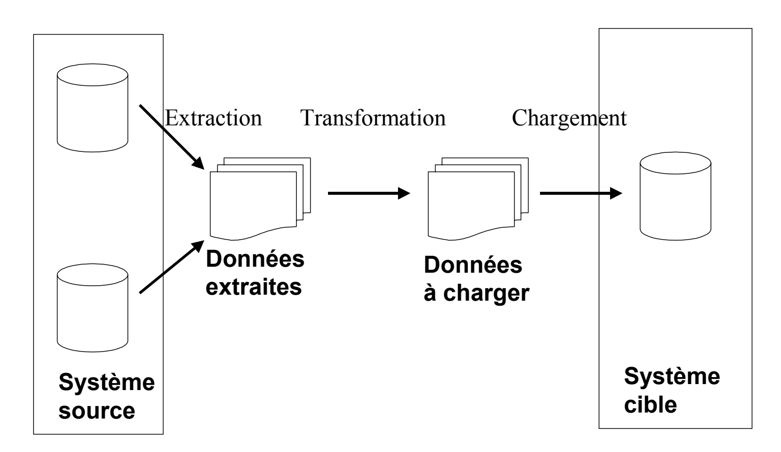


Figure 8.4. *Extraction, transformation et chargement des données*

8.4.1. *L'extraction des données*

Il s'agit d'extraire les données là où elles se trouvent. Les outils ETL ont la capacité de se connecter aux différentes bases de données ou fichiers des applications source. Des filtres permettent de ne sélectionner que les données pertinentes.

Si l'ETL assure la connexion avec les systèmes source, l'extraction proprement dite se fait généralement en utilisant les technologies suivantes :

– passerelles fournies par les éditeurs de SGBD,
– utilitaires de réplication et de mise à disposition de données,
– outils spécifiques.

8.4.2. *La transformation des données*

La transformation des données est la fonctionnalité principale de l'outil ETL. Elle permet de convertir une donnée issue du système source en une donnée adaptée au système cible.

On peut assimiler ce travail à celui d'un outil de traduction. L'ETL comprend donc les outils permettant de renseigner les règles de conversion. Ces outils peuvent être par exemple un langage ou une interface graphique de conversion.

Le travail de spécification des règles de transformation est essentiel, car il conditionne la qualité des données qui vont entrer dans le nouveau système.

8.4.3. *Le chargement des données*

C'est la dernière étape du processus. Une fois les données extraites et mises en phase avec la cible, les données sont chargées. L'ETL permet de contrôler les données avant chargement, pour s'assurer qu'elles sont en phase avec les règles de gestion.

8.4.4. *Quelques recommandations*

Parmi les nombreux fournisseurs de solutions ETL, citons Cognos, Business Objects, Ascential, Informatica, Sunopsis, Microsoft, Oracle… et il y en a bien d'autres.

Attention cependant, ces outils ne sont pas miraculeux. S'ils facilitent les tâches d'extraction du système source et de chargement dans le système cible, la charge de travail d'analyse et de mapping reste importante. Car l'outil doit être nourri des règles de transformation des données ce qui sous-entend un travail préalable de définition des règles de conversion. Ces règles pouvant être complexes, la maîtrise de l'ETL doit être parfaite et il est indispensable de bien mesurer le niveau de

compétence et de maîtrise de l'outil avant de se lancer dans son investissement et son utilisation.

Il peut en revanche être judicieux d'utiliser un ETL pour la reprise de données, si l'entreprise possède déjà un tel outil pour d'autres besoins. Si l'entreprise n'en possède pas, il peut aussi être intéressant d'en acquérir un, avec la perspective d'une utilisation future comme l'alimentation d'un système de Business Intelligence.

8.5. Conclusion

La migration d'applications informatiques nécessite de gérer la reprise des données des anciens systèmes vers les nouveaux. Une stratégie claire de reprise doit être établie, afin de garantir des données de qualité en phase avec les exigences des nouvelles applications.

La migration est l'occasion pour l'entreprise de nettoyer ses données pour repartir sur une base saine et tirer profit des possibilités offertes par la nouvelle application.

CHAPITRE 9

Facteurs-clés de succès de la reprise des données

Comme vu dans le chapitre précédent, la migration d'une application vers une autre nécessite une reprise des données afin d'assurer, d'une part, le maintien du patrimoine informationnel (clients, fournisseurs, articles), d'autre part, la continuité dans le travail opérationnel (commandes, ordres de production, en-cours comptables…).

Les opérations de reprise pouvant s'avérer complexes et compromettre le démarrage d'une nouvelle application, il est nécessaire de se doter de moyens adaptés pour garantir la réussite du projet.

9.1. Ressources et management

C'est incontestablement le point le plus important. Identifier et mettre en place les ressources adéquates pour la préparation de la reprise (développement d'outils, mapping) puis son exécution sont la clé de la réussite.

L'un des écueils les plus répandus est de considérer la reprise des données sous un angle purement technique. Or, si les outils

techniques de reprise sont incontournables, la qualité des données migrées reste le point majeur.

Il y a donc besoin d'associer à ce travail les *utilisateurs-clés du business* pour valider le contenu des fichiers et participer au mapping entre ancienne application et nouvelle application. Un responsable de la reprise des données assurant l'animation, la coordination et le suivi de la reprise doit être nommé dès le début du projet. Les ressources participant à la reprise des données doivent être correctement dimensionnées, sachant que le travail de préparation et de mapping est important.

Globalement, les intervenants sont les suivants :

– l'utilisateur-clé (*key user*) : sa connaissance de l'application actuelle ainsi que son implication dans la solution future, le positionne au premier plan des tâches de mapping et de validation des données à migrer ; c'est en quelque sorte le garant de la qualité des données qui seront chargées ; il y a généralement plusieurs key users, en fonction de la couverture fonctionnelle de la nouvelle application (comptabilité, contrôle de gestion, production…) ;

– le développeur d'outils de migration : que ce soit pour extraire les données ou les charger, la mise en place d'outils nécessite les connaissances techniques adéquates ; le ou les développeurs ont donc pour mission de développer les outils requis ; ils doivent être capables de dialoguer avec les équipes fonctionnelles ; si un outil ETL est utilisé, au moins un expert de l'outil doit être intégré à l'équipe ;

– l'expert fonctionnel : il travaille avant tout sur les spécifications ou le paramétrage de la future application ; sa connaissance de la cible fait qu'il participe également au travail de mapping ; comme pour le key user, il y a généralement plusieurs experts fonctionnels sollicités (au moins un par domaine fonctionnel) ;

– le responsable de la migration : c'est le chef d'orchestre de l'opération de migration ; il planifie, coordonne et suit le travail

de préparation des données ; lors de la bascule, il est évidemment « sur le pont » pour encadrer et animer les équipes participant à l'opération ; selon la taille du projet, ce peut être le chef de projet lui-même qui assure ce rôle.

Aussi, l'équipe de migration ainsi constituée doit être complètement intégrée à l'ensemble du projet. Les échanges entre les équipes fonctionnelles et l'équipe de migration doivent être réguliers.

9.2. Formation

La formation de l'ensemble des personnes participant au travail de préparation des données doit être organisée assez tôt. Il s'agit de bien expliquer le contenu de chaque donnée du système cible : définition, utilisation, codification. Plus cette formation sera approfondie, meilleure sera la qualité des données. Le travail de validation des données sera notamment plus simple si la connaissance précise des données cibles est partagée.

L'une des difficultés est d'organiser la formation au bon moment : il ne faut pas la prévoir trop tard. Pour autant, il ne faut pas la planifier trop tôt non plus, car les règles de gestion ne sont en général pas finalisées et risquent de changer. Pour pallier cette difficulté, la formation peut être organisée en deux temps :

– l'une après définition (et validation) des règles de gestion : elle concerne l'ensemble des intervenants de la migration ;

– l'autre après définition des formats de fichiers de chargement : elle concerne les personnes qui devront manipuler et valider ces fichiers avant chargement ; ce sont principalement les utilisateurs-clés.

9.3. Planification et suivi des tâches préparatoires à la bascule

La préparation de la bascule comprend principalement les tâches de :

– nettoyage des données,
– spécification des règles de gestion,
– formation,
– définition des règles de conversion,
– détermination du mode de reprise (automatique, manuel),
– définition de la séquence de reprise,
– développement d'outils de migration,
– définition des formats de fichiers de reprise,
– tests de la reprise,
– validation de la reprise.

L'ensemble de ces tâches fait l'objet d'un planning détaillé, généralement intégré au planning global du projet. Le suivi fin de ces tâches est effectué de façon quasi continue. Les dérapages éventuels doivent être décelés au plus tôt pour pouvoir réagir et mettre en place un plan d'actions pragmatique et énergique.

Insistons encore une fois sur le nettoyage des données : c'est clairement un facteur-clé de succès qui permet d'obtenir des données de qualité dans le système cible. L'élimination des doublons, de même que l'organisation d'inventaires physiques avant la bascule (si des stocks doivent être repris) sont des points-clés.

9.4. Tests

Que ce soit pour les données de base ou pour les données dynamiques, des tests sont évidemment nécessaires. Il est préconisé de dérouler des flux business avec des données migrées par les outils de reprise pour s'assurer, d'une part, que ces outils fonctionnent, d'autre part, que le mapping est correct.

C'est en déroulant des flux que l'on peut rapidement connaître le niveau de qualité des données reprises.

Pratiquement, les tests concernent toute la chaîne de migration : extraction, transformation et chargement. Là encore, les tests doivent être menés conjointement par les développeurs d'outils et les experts fonctionnels.

9.5. Validation des données

La validation des données doit être intégrée à la phase de tests de l'application. Il faut donc veiller à réaliser les tests sur des données réelles et extraites de l'ancienne application. Les données doivent être validées :

– avant chargement : les fichiers avant chargement permettent d'intervenir sur la chaîne de migration avant que les données ne soient chargées dans le système cible ; comme indiqué à la section 9.2, cela nécessite néanmoins que les personnes validant ces fichiers soient formées à la structure et au contenu de ces fichiers ;

– après chargement : une fois chargées, les utilisateurs-clés participant au projet ont la possibilité de vérifier le contenu de ce qui a été chargé dans la nouvelle application ; l'exécution de scénarios de tests simulant la réalité est certainement la meilleure façon de procéder.

9.6. Planification et suivi détaillé de la bascule

Un plan détaillé de bascule doit donc être élaboré et communiqué. Si les données de base (articles, clients, fournisseurs…) sont migrées à l'avance, il n'en est pas de même pour les données dynamiques. Comme précisé au chapitre 8, la bascule doit donc être orchestrée avec précision. Rappelons que la bascule a lieu pendant une période très courte, souvent le week-end, pour éviter de pénaliser le moins possible le travail opérationnel.

Le plan de bascule précise qui fait quoi, quand et où. Un plan de présence est diffusé. En France, il faut aussi prévoir d'informer les autorités administratives en cas de travail le week-end. Dans ce cas, la logistique associée doit être prévue : accès aux locaux, communications, réunions de suivi à heure précise, repas…

La réalisation de la bascule est souvent un moment fort. C'est l'aboutissement de plusieurs mois de travail qui se réalise dans un laps de temps très court. Les tâches de reprise des données occupent l'essentiel de la bascule.

Une fois la migration terminée et la nouvelle application démarrée, un suivi régulier doit être effectué avec une attention particulière à la qualité des données.

9.7. Conclusion

Plus qu'une opération technique, la reprise des données doit être considérée avant tout comme une opération fonctionnelle et « business ». D'où l'importance d'intégrer des personnes métier dans le processus de migration. Le succès d'une bascule repose principalement sur une bonne gestion des ressources participant à la bascule.

CONCLUSION

Vers une culture de l'information et des données

Nous l'avons vu au cours de cet ouvrage, la qualité des données joue un rôle majeur dans la performance et la compétitivité des entreprises.

D'énormes progrès restent à accomplir dans ce domaine et il n'est pas exagéré d'affirmer que nous sommes aux balbutiements de l'ère de l'information. Le monde évolue et l'information devient un capital de plus en plus précieux qu'il faut gérer comme une ressource stratégique. Il est intéressant de souligner que les données qui étaient d'une qualité exceptionnelle hier sont d'une qualité acceptable aujourd'hui et seront d'une qualité médiocre demain.

On peut clairement s'attendre à une accélération de la demande en données de qualité dans les années à venir. Aussi la variété de la demande va s'accentuer et l'une des difficultés sera précisément d'être capable de répondre à cette diversité. On s'oriente donc peu à peu vers un niveau d'exigence élevé avec une palette de besoins de plus en plus large.

Plus qu'une question technique, la recherche de la qualité des données doit être un état d'esprit, une vision partagée, une culture que l'entreprise doit développer pour sortir vainqueur de la bataille de l'information. La technologie (ordinateurs, bases de données, réseaux) est un support utile à l'optimisation de tâches bien définies. Dans ces conditions, elle améliore la productivité, réduit les coûts, libère les hommes et les femmes pour d'autres tâches. Cependant, son efficacité est moindre si le circuit d'information est à la base mal défini ou mal géré, car l'information sera certes produite plus rapidement, mais avec autant d'imperfections qu'auparavant. Les technologies de l'information seules ne permettent donc pas de faire des miracles. C'est pourquoi, avant de penser technologie, il faut réfléchir à la chaîne d'information.

Le succès dans la mise en place d'une politique d'entreprise centrée sur la qualité et la fluidité de l'information repose principalement sur la volonté des hommes et les femmes. L'adhésion des collaborateurs à un tel programme est évidemment indispensable et les dirigeants doivent clairement montrer la voie. Reconnaissons que des freins ne manqueront pas de se manifester pour des raisons diverses : peur du changement, crainte de perdre des responsabilités, voire son emploi, conflits de pouvoir, image négative du concept même de « qualité », etc.

C'est pourquoi l'évolution vers une culture d'entreprise laissant une place prépondérante à la qualité de l'information, doit être gérée avec une attention particulière aux questions humaines.

BIBLIOGRAPHIE

AKOKA J., *Conception des bases de données relationnelles*, Vuibert Informatique, Paris, 2001.

BERDUGO A., CORNIOU J.-P., MAHL R., PÉPIN J.-F., *Challenges pour les DSI : L'art du management des systèmes d'information en 21 études de cas*, Dunod, Paris, 2005.

DAUDIN J.-J., TAPIERO Ch., *Les outils et le Contrôle de la Qualité*, Economica, Paris, 1996.

DEIXONNE J.-L., *Piloter un projet ERP*, Dunod, Paris, 2003.

DELMOND M.-H., PETIT Y., GAUTIER J.-M., *Management des systèmes d'information*, Dunod, Paris, 2003.

DRUCKER P., *L'Avenir du Management*, Village Mondial, Paris, 1999.

ECKSTEIN R., *XML précis et concis*, O'Reilly France, Paris, 2000.

FENOULIÈRE Ph., *La qualité de l'informatisation*, Hermès, Paris, 1996.

FRANCO J.-M., LIGNEROLLES S. (de), *Piloter l'entreprise grâce au Data Warehouse*, Eyrolles, Paris, 2000.

GARDARIN G., *Bases de données : objet et relationnel*, Eyrolles, Paris, 2001.

GOGUE J.-M., *Management de la qualité*, Economica, Paris, 2001.

JEAN G., *Urbanisation du business et des SI*, Hermès, Paris, 2000.

JOLIOT D., *Management des SI*, Hermès, Paris, 2003.

LANGLOIS M., FAVERIO D., LESOURD M., *XML dans les échanges électroniques : le Framework ebXML*, Hermès, Paris, 2004.

LE FLOHIC G., *E-business : Guide de référence*, Elenbi, Paris, 2002.

LEFÉBURE R., VENTURI G., *Gestion de la relation clients*, Eyrolles, Paris, 2004.

LEJEUNE Y., *XML*, Editions Micro Application, Paris, 2004.

LEQUEUX J.-L., *Manager avec les ERP*, Editions d'Organisation, Paris, 2002.

MATHERON J.-P., *Comprendre Merise : Outils conceptuels et Organisationnels*, Eyrolles, Paris, 2002.

MINC A., NORA S., *L'informatisation de la société*, Le Seuil, Paris, 1978.

MOISAN D., *CRM, Gestion de la relation client*, Hermès, Paris, 2002.

OLSON J.E., *Data Quality : the accuracy dimension*, Morgan Kaufmann Publishers, San Francisco, 2002.

PETERS T., *Le prix de l'excellence*, Dunod, Paris, 1982.

RECHENMANN J.-J., *Internet & marketing*, Editions d'Organisation, Paris, 2001.

REDMAN T., *Data Quality For The Information Age*, Artech House Publishers, Londres, 1997.

REIX R., *Systèmes d'information et Management des organisations*, 5e édition, Vuibert Informatique, Paris, 2004.

SOUVAY P., *Statistique et qualité : Applications pratiques*, Association Française de Normalisation (AFNOR), Paris, 2002.

TARDIEU H., ROCHFELD A., COLLETTI R., *La Méthode Merise : Principes et outils*, Editions d'Organisation, Paris, 2000.

Sites internet

www.weblmi.com : site du Monde Informatique.

www.journalinformatique.com : actualités des technologies de l'information.

www.linternaute.com : magazine en ligne grand public de l'internet français.

www.01net.com : site du groupe Tests (01 Informatique, l'ordinateur individuel, etc.).

www.cigref.fr : site du Club informatique des grandes entreprises françaises.

www.searchcrm.com : site spécialisé dans la gestion de la relation client (CRM).

www.W3.org : site du World Wide Web Consortium.

… et tous les sites internet des fournisseurs de logiciels.

INDEX

Collection Management et Informatique

sous la direction de Nicolas Manson

Alain Amghar – Conduite opérationnelle des projets, 2004

Alain Berdugo – Guide du management des SI, 2001

Alain Berdugo – Le maître d’ouvrage du SI, 1997, 2005

Alain Berdugo et Pierre Aliphat – Systèmes informatiques de l'entreprise, 1997

Gilles Blanchard et André Vincent– La fonction achat en informatique et télécoms, 1999

Jean-Christophe Bonne et Aldo Maddaloni – Convaincre pour urbaniser le SI, 2004

Jean-Pierre Briffaut – Processus d'entreprise pour la gestion, 2004

Stéphane Calé – La gestion de projets de télécoms, 2005

Didier Caron – Méthodes objet, 1997

Jean-Pierre Corniou – La société de la connaissance, 2002

Denis Debaecker – PLM — La gestion collaborative du cycle de vie des produits, 2004

Bernard Debauche et Patrick Mégard – BPM, Business Process Management, 2004

Alain Desroches *et al.* – La gestion des risques, 2003

Tru Dô-Khac – Externalisation des télécoms d’entreprise, 2005

Luc Dorrer – Hommes et projets informatiques, 2004

Philippe Fenoulière – Vers une informatique ouverte, 2004

Philippe Fenoulière – La qualité de l’informatisation, 1996

Franck Franchin et Rodolphe Monnet – Le business de la cybercriminalité, 2005

Jean-François Gautier et Alan Fustec – Informatique de compétition, 1997

Thierry Harlé et Florent Skrabacz – Clés pour la sécurité des SI, 2004

Pierre Jaquet – Les réseaux et l’informatique d'entreprise, 2003

Gérard Jean – Urbanisation du business et des SI, 2000
Didier Joliot – Management des SI, 2003
Didier Joliot – Performances des SI, 2003
Henri Kloetzer – La maîtrise d'ouvrage des projets informatiques, 2002
Pierre Kraus – Prévision et maîtrise des performances d’un système informatique, 2005
Pierre Laigle – Dictionnaire de l’infogérance, 2000
Jean-Luc Lapon – La direction informatique et le pilotage de l’entreprise, 1999
Bernard Le Roux *et al.* – Urbanisation et modernisation du SI, 2004
Jean-Noël Lhuillier – Le management de l’information, 2005
Henry Ly – L’audit technique informatique, 2005
Pierre Maret – Ingénierie des savoir-faire, 1997
Jean-Pierre Meinadier – Le métier d'intégration de systèmes, 2002
Dominique Moisand – CRM, gestion de la relation client, 2002
Pascal Muckenhirn – Le SI décisionnel, 2003
Gilbert Nzeka – La protection des sites informatiques face au hacking, 2005
Jean-Louis Peaucelle – Informatique rentable et mesure des gains, 1997
Dany Provost – An 2000, la transition réussie, 1998
Luc Rubiello – Techniques innovantes en informatique, 1997
Alain Rugy (de) – Management et gestion du parc micro, 1998
Jacques Sassoon – Urbanisation des SI — épuisé, 1998
Pascal Silvestre – Le développement des SI, 1996
Marcel Soberman – Développement rapide d'applications, 1996
Sys-com – La bascule du SI vers l’euro – 2ème édition, 2000
Philippe Tassin – Systèmes d'information et management de crise, 2005
Marc Thorin – L’audit informatique, 2000

Zouheir Trabelsi et Henri Ly – La sécurité sur internet, 2005

Jean-Pierre Vickoff – Estimation et architecture des développements Agiles, 2005

Jean-Pierre Vickoff – Systèmes d'information et processus Agiles, 2003

COLLECTION ETUDES ET LOGICIELS INFORMATIQUES

sous la direction de Nicolas Manson

Alain Amghar et Frédéric Sitbon – Microsoft Office Project 2003, 2004

Yves Constantinidis – Le logiciel à valeur ajoutée, 2001

Yves Constantinidis – Outils de construction du logiciel, 1998

Erol Giraudy *et al.* – Le portail Microsoft Sharepoint, 2004

Guy Lapassat – Architecture fonctionnelle des logiciels, 2003

Guy Lapassat – Urbanisme informatique et architectures applicatives, 2003

Jean-Pierre Meinadier – Ingénierie et intégration des systèmes, 1998

Michel Priem – Trafic et performances des réseaux multi-services, 2004

Jacques Printz *et al.* – Coûts et durée des projets informatiques, 2001

Jacques Printz – Productivité des programmeurs, 2001

Jacques Printz – Puissance et limites des sytèmes informatisés, 1998

Yvon Rastetter – Le logiciel libre dans les entreprises, 2002

Marcel Soberman – Les grilles informatiques, 2003

Sys-com – Stratégie de test e-business, 2001

Spyros Xanthakis *et al.* – Le test des logiciels, 2000

COLLECTION NOUVELLES TECHNOLOGIES INFORMATIQUES

sous la direction de Nicolas Manson

Jean-Louis Bénard – Les portails d'entreprise, 2002

Jérôme Besancenot *et al.* – Les systèmes transactionnels, 1997

Christian Bonjean – Helpdesk, 1999

Jean-Pierre Briffaut – Systèmes d'information en gestion industrielle, 2000

Jean-François Goglin – La cohabitation électronique, 2005

Jean-François Goglin – La construction d'un datawarehouse – 2ème édition, 2001

Jean-François Goglin – Le Datawarehouse, pivot de la relation client, 2001

Jean-François Goglin et Philippe Usclade – Du client-serveur au web-serveur, 1999

Marc Langlois *et al.* – XML dans les échanges électroniques, 2004

Bernard Manouvrier – EAI, Intégration des applications d'entreprise, 2001

Norbert Paquel et Olivier Bezaut – XML et développement des EDI, 2002

Yvon Rastetter – La fusion de la téléphonie dans l'internet, 2005

René-Charles Tisseyre – Knowledge Management, 1999

Collection Synthèse Informatiques CNAM
sous la direction de Nicolas Manson

Florent Bastianello – L'informatique, mémoire de l'entreprise, 1996
Jean-André Biancolin – Temps réel, 1995
Bruno Dardonville – Architecture de Windows NT, 1996
Hervé Guitter – La compression des images numériques, 1995
Renaud Hilleret – Stockage des données distribuées, 1996
Jean-Marc Lê – Les systèmes de télécoms mobiles, 1998
Lionel Mallordy – Répartition d'objets dans les bases de données, 1995
Christophe Pasquier – L'approche objet, 1995
Catherine Pérou – La dépense informatique en France, 1996
Ricardo Ruiz – Systèmes de gestion de bases de données orientés objet, 1997
Laurent Schneider – Logiciels serveurs et outils d'administration pour le web, 1997
Philippe Vetter – Calcul coopérant par passage de messages, 1995
Bernard Weiss – ATM, 1995
Bernard Zignin – Les techniques du multithread, 1996